```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:      June 11, 2003, 20:06:42 ; Search time 2303 Seconds
                                        (without alignments)
                                        3866.895 Million cell updates/sec

Title:       US-09-308-397-2
Perfect score: 1518
Sequence:    1 MTKTAFLPAGQGAQYLGMGR.........QTAHLAHVEDQASLVALLEK 306

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:    2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters:    4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool/US09308897/runat_06062003_112348_28873/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09308397 @CGN_1_1652 @runat_06062003_112348_28873 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   GenEmbl:*
        1: gb_ba:*          29: em_vi:*
        2: gb_htg:*         30: em_htg_hum:*
        3: gb_in:*          31: em_htg_inv:*
        4: gb_om:*          32: em_htg_other:*
        5: gb_ov:*          33: em_htg_mus:*
        6: gb_pat:*         34: em_htg_pln:*
        7: gb_ph:*          35: em_htg_rod:*
        8: gb_pl:*          36: em_htg_mam:*
        9: gb_pr:*          37: em_htg_vrt:*
       10: gb_ro:*          38: em_sy:*
       11: gb_sts:*         39: em_htgo_hum:*
       12: gb_sy:*          40: em_htgo_mus:*
       13: gb_un:*          41: em_htgo_other:*
       14: gb_vi:*
       15: em_ba:*
       16: em_fun:*
       17: em_hum:*
       18: em_in:*
       19: em_mu:*
       20: em_om:*
       21: em_or:*
       22: em_ov:*
       23: em_pat:*
       24: em_ph:*
       25: em_pl:*
       26: em_ro:*
       27: em_sts:*
       28: em_un:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1518	100.0	921	6	AX194049	AX194049 Sequence
	2	1518	100.0	11523	1	AF197933	AF197933 Streptoco
	3	1518	100.0	12039	1	AE007354	AE007354 Streptoco
c	4	1518	100.0	19702	6	BD003687	BD003687 Polynucle
	5	1518	100.0	111135	2	SPNEU01906	AL449928 Streptoco
	6	1515	99.8	945	6	AX194275	AX194275 Sequence
	7	1515	99.8	10925	1	AE008418	AE008418 Streptoco
	8	1048	69.0	10701	1	AE006311	AE006311 Lactococc
	9	1036	68.2	10909	1	AE006603	AE006603 Streptoco
c	10	1036	68.2	50416	1	AE014165	AE014165 Streptoco
c	11	1030	67.9	14456	1	AE010088	AE010088 Streptoco
c	12	858	56.5	1196	6	AR193756	AR193756 Sequence
	13	775.5	51.1	11845	1	AE013105	AE013105 Thermoana
c	14	710.5	46.8	13165	1	AE007854	AE007854 Clostridi
c	15	703.5	46.3	298050	1	AP003189	AP003189 Clostridi
	16	669.5	44.1	27779	1	BSY13937	Y13937 Bacillus su
	17	669.5	44.1	208780	1	BSUB0009	Z99112 Bacillus su
	18	657.5	43.3	14530	1	AE004276	AE004276 Vibrio ch
c	19	654	43.1	303249	1	AP001515	AP001515 Bacillus
	20	643.5	42.4	1440	1	ECFABHDG	Z11565 E.coli fabH
	21	643.5	42.4	1608	1	ECOFABD	M87040 Escherichia
	22	643.5	42.4	13051	1	AE000210	AE000210 Escherichi
	23	643.5	42.4	16379	1	D90745	D90745 Escherichia
	24	640.5	42.2	12890	1	AE005319	AE005319 Escherich
	25	640.5	42.2	222605	1	AP002555	AP002555 Escherich
	26	636.5	41.9	2678	1	BSU59433	U59433 Bacillus su
c	27	636	41.9	10857	1	U32701	U32701 Haemophilus
	28	628.5	41.4	113050	1	AE013778	AE013778 Yersinia
	29	628.5	41.4	193050	1	AJ414149	AJ414149 Yersinia
	30	627.5	41.3	950	6	AX432728	AX432728 Sequence
	31	627.5	41.3	960	6	AX432671	AX432671 Sequence
	32	626	41.2	10684	1	AE006228	AE006228 Pasteurel
	33	616.5	40.6	4430	1	AF044668	AF044668 Salmonell
	34	616.5	40.6	20097	1	AE008752	AE008752 Salmonell
	35	612.5	40.3	254050	1	AL627269	AL627269 Salmonell
	36	611.5	40.3	6394	1	AB021978	AB021978 Vibrio ma
	37	598	39.4	193050	1	AL646062	AL646062 Ralstonia
	38	594.5	39.2	14103	1	AE000712	AE000712 Aquifex a
	39	589.5	38.8	2301	6	AX416528	AX416528 Sequence
c	40	588.5	38.8	347050	1	AL591981	AL591981 Listeria
c	41	586	38.6	313450	1	AL596170	AL596170 Listeria
	42	586	38.6	319630	6	AX413016	AX413016 Sequence
	43	586	38.6	349980	6	AX417046	AX417046 Sequence
c	44	578	38.1	9922	1	AE012960	AE012960 Chlorobiu
	45	572.5	37.7	11361	1	AE011741	AE011741 Xanthomon

ALIGNMENTS

RESULT 1

```
AX194049
LOCUS       AX194049               921 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 30 from Patent WO0149721.
ACCESSION   AX194049
VERSION     AX194049.1  GI:15211644
KEYWORDS
SOURCE      Streptococcus pneumoniae.
ORGANISM    Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 921)
AUTHORS     Dougherty,T.J., Pucci,M.J., Dougherty,B.A., Davison,D.B.,
            Bruccoleri,R.E. and Thanassi,J.A.
TITLE       Novel bacterial genes and proteins that are essential for cell
            viability and their uses
JOURNAL     Patent: WO 0149721-A 30 12-JUL-2001;
            Bristol-Myers Squibb Co. (US)
FEATURES             Location/Qualifiers
     source          1..921
                     /organism="Streptococcus pneumoniae"
                     /db_xref="taxon:1313"
BASE COUNT      246 a    193 c    232 g    250 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                    1.44e-109           Length:        921
Percent Similarity:       1518.00             Matches:       306
Best Local Similarity:    100.00%             Conservative:    0
Query Match:              100.00%             Mismatches:      0
DB:                       100.00%             Indels:          0
                          6                   Gaps:            0

US-09-308-397-2 (1-306) x AX194049 (1-921)

Qy   1  MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg   20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1  ATGACTAAAAACAGCCCTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG   60

Qy  21  AspPheTyrAspLeuValLysProIleValGlyThrIleAspArgAlaSerGlnValLeu   40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61  GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGTCAGGTGCTC  120

Qy  41  GlyTyrAspLeuArgTyrLeuIleAspThrGluGluLysLysLeuAsnGlnThrArgTyr   60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121  GGTTATGATTTGCGTTATCTCATCGATACGGAAGAAGAGAAACTCAATCAGACCCGCTAT  180

Qy  61  ThrGlnProAlaIleLeuAlaThrSerValAlaIleThrTyrArgLeuLeuGlnLysGly   80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181  ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC  240

Qy  81  TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer  100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241  TATCAGCCTGATATGGTCGCTGGTTTGTCTCTTGGAGAATACTCTGCCTTGGTGGCAAGC  300

Qy 101  GlyAlaLeuAspPheGluAspPheAsnAspGluAlaAlaLeuValAlaLysGlyAlaTyrMetGlu  120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301  GGCGCCTTGGATTTTGAAGATGCGGTTGCCTGGCTAGCTAAGCGGGAGGCCTATATGGAA  360

Qy 121  GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal  140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361  GAAGCCGCTCCTGCAGACTCGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC  420

Qy 141  IleGluGluAlaCysGlnLysAlaSerGlnLysAlaSerGlyLysValThrProAlaAsnTyrAsn  160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421  ATTGGAGGAAGGTCAAAAAGCTTCTGGAGTGGTTACTCCCAGCCAACTATAAC  480

Qy 161  ThrProAlaGlnIleValIleGluAlaGlyValGluValAlaValAspArgArgAlaValGluLeu  180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481  ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGATCGAGCGGTTGAACTT  540

Qy 181  LeuGlnGluAlaGlyValAlaLysArgLeuIleProLeuValLysValSerGlyValProPheHisThr  200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541  TTGCAAGAAGCAGGTGTCGCAAACGCTTAATCCTCTTAAGGTGTCAGGTGTCCCTTTCACACC  600
```

```
Qy 201  AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer  220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601  GCTCTCCTTGAGCCAGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTTCA  660

Qy 221  AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGlnGluAspIle  240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661  GATTTTACTTGTCCCTTCGGCAATACAGAAGCTGCTGATGCAAAAAGAGGACATT  720

Qy 241  AlaGlnLeuLeuThrArgArgGlnValLysGlyValArgPheTyrGluSerIleGlyValVal  260
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721  GCTCAGCTCTTGACGCGTCAGGTCAAGGAACCCGTTCGTTCTTCATGAGAAGTATTGGGGTC  780

Qy 261  MetGlnGluAlaGlyIleLeuSerAsnPheIleGluIleGlyGlyLysProGlyLysLeuSerGly  280
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781  ATGCAAGAAGCAGGCATAAGCAACTTTATCGAGATTGGACCGGGAAAGTCTGTCAGGT  840

Qy 281  PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGlnSerGlnAlaSerLeu  300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841  TTTGTTAAAAAAATTGATCAAACGCTCACTTAGCTCATGTGGAAGATCAAGCCAGTTTA  900

Qy 301  ValAlaLeuLeuGluLys  306
        ||||||||||||||||||
Db 901  GTAGCACTTTTAGAAAAA  918
```

```
RESULT 2
AF197933
LOCUS       AF197933              11523 bp    DNA     linear   BCT 10-AUG-2000
DEFINITION  Streptococcus pneumoniae fab gene cluster, complete sequence.
ACCESSION   AF197933
VERSION     AF197933.1  GI:9789228
KEYWORDS
SOURCE      Streptococcus pneumoniae.
ORGANISM    Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 11523)
AUTHORS     Heath,R.J. and Rock,C.O.
TITLE       A triclosan-resistant bacterial enzyme
JOURNAL     Nature 406 (6792), 145-146 (2000)
MEDLINE     20365714
PUBMED      10910344
REFERENCE   2  (bases 1 to 11523)
AUTHORS     Heath,R.J. and Rock,C.O.
TITLE       Direct Submission
JOURNAL     Submitted (25-OCT-1999) Department of Biochemistry, St Jude
            Children's Research Hospital, 332 North Lauderdale Street, Memphis,
            TN 38105, USA
FEATURES             Location/Qualifiers
     source          1..11523
                     /organism="Streptococcus pneumoniae"
                     /strain="R6"
                     /db_xref="taxon:1313"
     gene            210..1184
                     /gene="fabH"
     CDS             210..1184
                     /gene="fabH"
                     /note="FabH; KASIII; acetoacetyl-ACP synthase; by
                     similarity to FABH_ECOLI (39% identity)"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-ketoacyl-ACP synthase III"
                     /protein_id="AAF98271.1"
                     /db_xref="GI:9789229"
                     /translation="MAFAKISQVAHYVPEQVTNHDLAQIMDTNDEWISRTGIRQRH
                     ISRTESTSDLATEVAKKLMAKAGITGEELDFIILAIITPDSMPSTAARVQANIGANK
                     AFAFDLTAACSGFVFALSTAEKFIASGRFQKGLVIGSETLSKAVDWDSRSTAVLFGDG
                     AGGVLEASEQEHFLAEFSLNSDGSRECLIYGHSGLHSPFSDQESADSFLKMDGRTVF
                     DFAIRDVAKSIKQTIDESPIEVTDLDYLLLHQANDRILDKVNARKIGVDRAKLPANMME
                     YGNTSAASIPILLSEVEQGLIPLDGSQTVLLSGFGGGLWGTLILTI"
     gene            1244..1468
                     /gene="acpP"
     CDS             1244..1468
```

```
          /gene="acpP"
          /note="ACP; by similarity to ACPP_ECOLI (38% identity)"
          /codon_start=1
          /transl_table=11
          /product="acyl carrier protein"
          /protein_id="AAF98272.1"
          /db_xref="GI:9789230"
          /translation="MAVFEKVQEIIVEELGKDASEVTLESTFDDLDADSLDLFQVISE
          IEDAFDIQIEAERNDLKTVGDLVAYVEEQAK"
gene      1587..2561
          /gene="fabK"
CDS       1587..2561
          /gene="fabK"
          /note="FabK; NADH-dependent, FAD-containing enoyl-ACP
          reductase"
          /codon_start=1
          /evidence=experimental
          /transl_table=11
          /product="trans-2-enoyl-ACP reductase II"
          /protein_id="AAF98273.1"
          /db_xref="GI:9789231"
          /translation="MKTRITELLKIDYPIFQQGMAWVADQDLAGAVSKAGLGLIGGG
          NAPKEVVKANIDKIKSLIDKPGVNIMLLSPPVEDIVDLVIEEGVKVVTTGAGNPSKY
          MERFHEAGIIVIPVVPSVALAKMEKIGADAVIAEGMEAGHIGKJTTMTLVRQVATA
          ISIPVIRAGGIADGEGAAAGFMLGAEAVQVGFRFVAKESNAHPNYKEKILKARDIDT
          TISAQHFGHAVRAIKNQLTRDFELAEKDAFKQEDPDLEIFEQMGAGALAKAVVHGGVD
          GGSVMAQCIAGLVSKEETAEEILKDLYYGAAKKIQEEASRNTGVVRND"
gene      2554..3474
          /gene="fabD"
CDS       2554..3474
          /gene="fabD"
          /note="FabD; by similarity to FABD_ECOLI (42% identity);
          MCAT"
          /codon_start=1
          /transl_table=11
          /product="malonyl-CoA:ACP transacylase"
          /protein_id="AAF98274.1"
          /db_xref="GI:9789232"
          /translation="MTKTAFLFAGGQAQYLGMGRDFYDQYPIVKETIDRASQVLGYDL
          RYLIDTEDKLNQTRYTQPAILATSVAIYRLLIQEKGYQPDMVAGLSLGEYSALVASGA
          LDFEDAVALVAKRGAYMEEAAPADSGKMVAVLNTPVEVIEEACQKASELGVVTPANYN
          TPAQIVIAGEVVAVDRAVELLQEAGAKRLIPLKVSGPFHTALLEPASQKLAETLAQVS
          FSDPTCPIVGNTAAAVMQKED:AQLLTRQVKSPVRFYESIGVMQEAGISNFIEIGPGK
          VLSGFVKKIDQTAHLAHVEDQASLVALLEK"
gene      3508..4239
          /gene="fabG"
CDS       3508..4239
          /gene="fabG"
          /note="FabG; by similarity to FABG_ECOLI (44% identity)"
          /codon_start=1
          /transl_table=11
          /product="beta-ketoacyl-ACP reductase"
          /protein_id="AAF98275.1"
          /db_xref="GI:9789233"
          /translation="MKLEHKNIFITGSSRGIGLAIAHKFAQAGANIVLNSRGAISEEL
          LAEFSNYGIKVVPISGDVSDFADAKRMIDQAIAELGSVDVLVNNAGITQDTLMLKMTE
          ADFEKVLKVNLTGAFNMTQSVLKPMMKAREGAIINMSSVVGLMGNIQOANYAASKAGL
          IGFTKSVAREVASRNIRVNVIAPGMIESDMTAILSDKIKEATLAQIPMKEFGQAEQVA
          DLTVFLAGQDYITGQVVAIDGGLSM"
gene      4261..5496
          /gene="fabF"
CDS       4261..5496
          /gene="fabF"
          /note="FabF; KASII; oxoacyl-ACP synthase II; by similarity
          to FABF_ECOLI (43% identity) and FABB_ECOLI (33%
          identity)"
          /codon_start=1
          /transl_table=11
          /product="beta-ketoacyl-ACP synthase II"
          /protein_id="AAF98276.1"
          /db_xref="GI:9789234"
          /translation="MKLNRVVVTGYGVTSPIGNTPEEFWNSLATGKIGIGGITKFDHS
          DFDVHNRAAEIQDFPFDKYFVKKDTNRFDNYSLYALYAAQEAVNHANLDVEALNRDRFG

          VIVASGIGGIKEIEDQVLRLHEKGPKRVKPMTLPKALPNMASGNVAMRFGANGVCKSI
          NTACSSNDAIGDAFRSIKFGFQDVMLVGGTEASITPAIAGFQALTALSTTEDPTRA
          SIPEDKDRNGFVMGEGSGMLVLESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEG
          QGAIKAIKLALEEAEISPEQVAYVNAHGTSTPANEKGESGAIVAVLGKEVPVSSTKSF
          TGHLLGAAGVEAIVTIEAMRHNFPVMTAGTSEVSDYIEANVVYQGGLEKEIPYAISN
          TFGFGGHNAVLAFKRWENR"
gene      5499..5984
          /gene="accB"
CDS       5499..5984
          /gene="accB"
          /note="BCCP; by similarity to BCCP_ECOLI (37% identity);
          subunit of acetyl-CoA carboxylase"
          /codon_start=1
          /transl_table=11
          /product="biotin carboxyl carrier protein"
          /protein_id="AAF98277.1"
          /db_xref="GI:9789235"
          /translation="MNLNDIKDLMTQFDQSSLREFSYKNGTDELQFSKNEARPVPEVA
          TQVAPAVLATPSPVAPTSAPAETVAEEVPAPAEASVATEGNLVESPLVGVVYLAAGP
          DKPAFVTVGDSVKKGQTLVIIEAMKVMNEIPAPKDGVVTEILVSNEEMVEFGKGLVRI
          K"
gene      5981..6403
          /gene="fabZ"
CDS       5981..6403
          /gene="fabZ"
          /note="FabZ; by similarity to FABZ_ECOLI (46% identity)"
          /codon_start=1
          /transl_table=11
          /product="beta-hydroxyacyl-ACP dehydratase"
          /protein_id="AAF98278.1"
          /db_xref="GI:9789236"
          /translation="MIDIQGIKEALPHRYMLLVDRVLEVSEDTIVAIKNVTINEPFF
          NGHFPQYPVMPGVVIMEALAQTAGVLELSKPENKGKLVFYAGMDKVKFKQVVPGDQL
          VMTATFVKRRGTIAVVEAKAEVDGKLAASGTLTFAIGN"
gene      6415..7782
          /gene="accC"
CDS       6415..7782
          /gene="accC"
          /note="AccC; by similarity to ACCC_ECOLI (53% identity)"
          /codon_start=1
          /transl_table=11
          /product="acetyl-CoA carboxylase biotin carboxylase
          subunit"
          /protein_id="AAF98279.1"
          /db_xref="GI:9789237"
          /translation="MFRKILIANRGEIAVRIIRAARELGIATVAVYSTADKEALHTLL
          ADEAVCIGPGKATESYLNINAVLSAAVLTBEAAIHPGFGFLSENSKFATMCEEVGIKF
          IGPSGHVMDMMGDKINARAQMIKAGVPVIPGSDGEVHNSEEALIVAEKIGYPVMLKAS
          AGGGGKGIRKVEKPDDLVSAFETASSEAKANYGNGAMYIERVIYPARHIEVQILGDEH
          GHVIHLGERDCSLQRNNQKVLEESPSIAIGKTLRHEIGAAAVRAAEFVGYENAGTIEF
          LLDEASNFYFMEMNTRVQVHPVTFEVSGVDIVKEQICIAAGQPLSVKQEDIVLRGH
          AIECRINAENPAFNPAPSPGKITNLYLPSGGVGLRVDSAVYPGYTIPPYDSMIAKII
          VHGENRFDALMKMQRALYELEIEGVQTNADFQLDLISDRNVIAGDYDTSFLMETFLPK
          YQEKE"
gene      7819..8685
          /gene="accD"
CDS       7819..8685
          /gene="accD"
          /note="AccD; by similarity to ACCD_ECOLI (43% identity)"
          /codon_start=1
          /transl_table=11
          /product="acetyl-CoA carboxylase beta subunit"
          /protein_id="AAF98280.1"
          /db_xref="GI:9789238"
          /translation="MALFSKDKYIRINPNRSVREKPQAKPEVPDELFSQCPGCKHTI
          YQKDLGSERICPHCSYTFRISAQERLALTIDMGTFKELFTGIESKDPLHFPGYQKKLA
          SMREKTGLHEAVVTGTALIKGQTVALGIMDSNFIMASMGTVVGEKITRLFEYATVEKL
          PVVLFTASGGARMQEGIMSLMQMAKISAAVKRHSNAGLFYLTILTDPTTGGVTASPAM
          EGDIILABPQSLVGPAGRRVIENTVRESLPEDFQKAEFLLEHGFVDAIVKRRDLPDTI
          ASLVRLHGGSPR"
gene      8682..9449
          /gene="accA"
CDS       8682..9449
```

```
Db     3334  ATGCAAGAAGCAGGCATAAGCAACTTTATCGAAGATGGACCGGGGAAAGTCTTGTCAGGT  3393
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      281  PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu  300

Db     3394  TTTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA  3453
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      301  ValAlaLeuGluLys  306

Db     3454  GTAGCACTTTTAGAAAAA  3471
```

Alignment Scores:
Pred. No.:   3.62e-108
Score: 1518.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB:

RESULT 3
AE007354
LOCUS   AE007354   12039 bp   DNA   linear   BCT 31-AUG-2001
DEFINITION Streptococcus pneumoniae TIGR4 section 37 of 194 of the complete
            genome.
ACCESSION   AE007354 AE005672
VERSION   AE007354.1 GI:11491888
KEYWORDS   .
SOURCE   Streptococcus pneumoniae TIGR4.
  ORGANISM Streptococcus pneumoniae TIGR4
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 12039)
  AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
            Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L.,
            McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
            Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
            Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
            Fraser,C.M.
  TITLE   Complete genome sequence of a virulent isolate of Streptococcus
            pneumoniae
  JOURNAL   Science 293 (5529), 498-506 (2001)
   MEDLINE   21357209
   PUBMED   11463916
REFERENCE   2 (bases 1 to 12039)
  AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
            Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A.,
            Feldblyum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,
            Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
            Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES   Location/Qualifiers
     source   1..12039
                     /organism="Streptococcus pneumoniae TIGR4"
                     /strain="TIGR4"
                     /db_xref="taxon:170187"
     gene   90..1064
                     /gene="SP0419"
     CDS   90..1064
                     /gene="SP0419"
                     /codon_start=1
                     /transl_table=11
                     /product="enoyl-(acyl-carrier-protein) reductase"
                     /protein_id="AAK74582.1"
                     /db_xref="GI:14971889"
                     /translation="MKTRITBLLKIDYPIFQGGMAWVADDGLAGAVSKAGGLGIIGGG
                     NAPKEVVKANIDKIKSLTDKPFGVNIMLLSPFVEDIVDLVIEEGVXVVTTGAGNPSKY
                     MERFHEAGIIVIPVVPSVALAKRMEKIGADAVIEGMEAGHIGKLTTMTLVRQVATA
                     ISIPVIAAGGIADGEAAGMLGAEAVQVTTRFVVAKESNAHPNYKEKILKARDIDT
                     TISAQHFGHAVRAIKNQLTRDFELAEKDAFKQEDPLEIFMGAGALAKAVVHGDVD
                     GGSVMAGQIAGLVSKEETABEILKDLYYGAAKKIQEEASRWTGVVRND"
     gene   1057..1977
                     /gene="SP0420"
     CDS   1057..1977
                     /gene="SP0420"
```

```
US-09-308-397-2 (1-306) x AF197933 (1-11523)

/gene="accA"
/note="AccA; by similarity to ACCA_ECOLI (46% identity)"
/codon_start=1
/transl_table=11
/product="acetyl-CoA carboxylase alpha subunit"
/protein_id="AAF98281.1"
/db_xref="GI:9789239"
/translation="MNIAKIVRBAREQSRLTTLDFATGIFDEFIQLHGDRSFRDDGAV"

Length:      11523
Matches:     306
Conservative:  0
Mismatches:   0
Indels:       0
Gaps:         1

Qy       1  MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2554  ATGACTAAAAACAGCCTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG  2613

Qy      21  AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2614  GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCCAGGTGCTC  2673

Qy      41  GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr  60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2674  GGTTATGATTTGCGTTATCTCATCGATACGGAAGAAGACAAAACTCAATCAGACCCGCTAT  2733

Qy      61  ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLeuLysGly  80
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2734  ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTAICTACCGTTTATTGCAAGAAAAAGGC  2793

Qy      81  TyrGlnProAspMetValAlaGlyValLeuSerLeuGlyValGluTyrSerAlaLeuAlaSer  100
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2794  TATCAGCCTGATATGGTTGCGGTTTGTCTTCTTGGAGAATACCTGCTTGGTGGCAAGT  2853

Qy     101  GlyAlaLeuAspPheGluGluAspLysValAlaAlaLeuValAlaLysArgGlyAlaTyrMetGlu  120
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2854  GGCGCCCTTGGATTTGAAGATGCGGTTGCCTTGGTAGCTAAGCGTGGAGCCTATATGGAA  2913

Qy     121  GluAlaAlaProAlaAspSerGlyLysMetValAlaLeuAsnThrProValGluVal  140
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2914  GAAGCCGGCTCCTGCTACCTCGGCAAGATGGTAGCCAGTTCTCAATACGCCAGTAGAGGTC  2973

Qy     141  IleGluGluGlnAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn  160
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2974  ATTGAAGAACAGGCCTGTCAAAAAGCTTCTGAACTTGGAGTGGTTACTCCAGCCAACTATAAC  3033

Qy     161  ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValAlaGluLeu  180
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3034  ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGATCGAGCGGTTGAACTT  3093

Qy     181  LeuGlnGluAlaGlyAlaGlyAlaLysAlaLysArgLeuIleSerGlyValSerGlyProPheHisThr  200
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3094  TTGCAAGAAGCAGGTGCCAAAGCTGCCAAACGCTTAATCTCTTTAAGGTGTGCAGGTGTCCCTTTCACACC  3153

Qy     201  AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer  220
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3154  GCTCTCCTTGAACCTGCTAGCCAGAAACTAGCTGAAACTCTAGCCCAGGTAAGTTTTTCA  3213

Qy     221  AspPheThrCysProLeuValGlyTyrAsnThrGluAlaValMetGlnLysGluAspIle  240
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3214  GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGCTGTGATGCAAAAAGAGGACATT  3273

Qy     241  AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal  260
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3274  GCTCAGCTCTTGACGCGTCAGGTCAAGGAACCCGTTCGTTTCTATGAAAGTATTGGGGTC  3333

Qy     261  MetGlnGlnAlaGlyIleTyrIleSerArgAsnPheIleGluIleGlyProGlyLysValLeuSerGly  280
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
                    /note="identified by match to TIGR    protein family HMM
                    TIGR00128"
                    /codon_start=1
                    /transl_table=11
                    /product="malonyl CoA-acyl carrier protein transacylase"
                    /protein_id="AAK74583.1"
                    /db_xref="GI:14971890"
                    /translation="MTKTRFLFAGGGAQYLGMGRDFYDCYPIVKETIDRASQVLGYDL
                    RYLIDTEEDKLNQTRYTQPAILATSVAITRLLQEKGYQPDMVAGLSLGEYSALVASGA
                    LDFEDAVALVAKRGAYMEEAAPADSGKMVAVLNTPVEVIEEACQKASELGVVTPANYN
                    TPAQIVRAAKGAYKGAVAVDRAVELLQEAGAKRLIPLKVSGPFHTALLEPASQKLAETLAQVS
                    FSDFTCPLVGNTEAAVMQKEDIAQLLTRQVKEPVRFYESIGVMQEAGISNFIEIGPGK
                    VLSGPVKKIDQTAHLAHVEDQASLVALLEK"
gene            2011..2742
                    /gene="SP0421"
CDS             2011..2742
                    /gene="SP0421"
                    /codon_start=1
                    /transl_table=11
                    /product="3-oxoacyl-[acyl-carrier protein] reductase"
                    /protein_id="AAK74584.1"
                    /db_xref="GI:14971891"
                    /translation="MKLEHKNIFITGSSRGIGLAIAHKFAQAGANIVLNSRGAISEEL
                    LAEPSNYGIKVVPISGDVSDFADAKRMIDQAIAELGSVDVLVNNAGITQDTLMLKMTE
                    ADFBKVLKVNLTGARNMTQSVLKPMMKAREGAIINMSSVVGLMGNIGQANYAASKAGL
                    IGFTKSVARFVASRKIRVNVIAPGMESDMTAILSDKIKEATLAQIPMKEFGQAEQVA
                    DLTVFLAGGDYLTGQVVAIDGGLSM"
gene            2764..3999
                    /gene="SP0422"
CDS             2764..3999
                    /gene="SP0422"
                    /codon_start=1
                    /transl_table=11
                    /product="3-oxoacyl-(acyl-carrier-protein) synthase II"
                    /protein_id="AAK74585.1"
                    /db_xref="GI:14971892"
                    /translation="MKLNRVVVTGVGVTSPIGNTPEEFWNSLATGKIGIGGITKFDHS
                    DFDVHNAAEIQDFPFDKYFVKKDTNRFDNYSLYAILYAAQEAVNHANLDVEALNRDRFG
                    VIVASGIGGIKEIEDQVLRLHEKGPKRVKPMTLPKALPNMASGNVAMRFGANGVCKSI
                    NTACSSNDAIGDAFRSIKFGFQDVMLVGGTEASITPFAIAGFQALTALSTTEDPTRA
                    SIPFDKDRNGFVMGEGSGMLVLESLEHAEKRGATILAEVVGYGNTCDAYHMTSPHPEG
                    QGAIKAIKLALEEAEISPEQVAYVNAHGTSTPANEKGESGAIVALVGKEVPVSSTKSF
                    TGHLLGAAGAVEAIVTIEAMRHNFVPMTAGTSEVSDYIEANVVYGGQGLEKEIPYAISN
                    TFGFGGHNAVLAFKKWENR"
gene            4002..4487
                    /gene="SP0423"
CDS             4002..4487
                    /gene="SP0423"
                    /note="identified by match to TIGR    protein family HMM
                    TIGR00531"
                    /codon_start=1
                    /transl_table=11
                    /product="acetyl-CoA carboxylase, bitoin carboxyl carrier
                    protein"
                    /protein_id="AAK74586.1"
                    /db_xref="GI:14971893"
                    /translation="MNLNDIKDLMTQFDQSSLREFSYKNGTDELQFSKNEARPVEVA
                    TQVAPAPVLATPSPVAPTSAPAETVAEEVPAPAEASVATEGNLVESPLVGVVVLAAGP
                    DKPAFVTVGDSVKKGQTLVIIEAMKVMENIPAPKDGVVTEILVSNEEMVEFGKGLVRI
                    K"
gene            4484..4906
                    /gene="SP0424"
CDS             4484..4906
                    /gene="SP0424"
                    /codon_start=1
                    /transl_table=11
                    /product="(3R)-hydroxymyristoyl-(acyl-carrier-protein)
                    dehydratase"
                    /protein_id="AAK74587.1"
                    /db_xref="GI:14971894"
                    /translation="MIDIQGVIKEALPHRYPMLLVDRVLEVSEDTIVAIKNVTINEPFF
                    NGHFQYPVMPGVVIMEALAQTAGVLELSKPENKGKLVFYAGMDKVKFKQVVPGDQL
                    VMTATFVKRGTIAVVEAKAEVDGKLAASGTLTFAIGN"
```

```
gene            4918..6285
                    /gene="SP0425"
CDS             4918..6285
                    /gene="SP0425"
                    /note="identified by match to PFAM    protein family HMM
                    PF00289"
                    /codon_start=1
                    /transl_table=1
                    /product="acetyl-CoA carboxylase, biotin carboxylase"
                    /protein_id="AAK74588.1"
                    /db_xref="GI:14971895"
                    /translation="MFRKILIANRGEIAVRIIRAARELGIATVAVYSTADKEALHTLL
                    ADEAVCIGPGKATESYLNINAVLSAAVLTEREAIHPGFGFLSENSKFATMCEVGIKF
                    IGPSGHVMDMGKINARAQMIKAGVPVIPGSDGEVHNSEEALIVAEKIGYPVMLKAS
                    AGGGKGIRKVEKPDDLVSAFETASEAKANYGNGAMYIERVIYPARHIEVQILGDEH
                    GHVIHLGRDCSLQRNQKVLESPSIAIGKTLRHEIGAAAVRAAEFVGYENAGTIEF
                    LLDEASSNFYFMEMNTRVQVEHFVTEFVSGVDIVKEQICIAAGQPLSVKQEDIVLRGH
                    AIECRINAENPAFNFAPSPGKITNLLPSGVGLRVDSAVPGYTIPPYYDSMIAKII
                    VHGENRFDALMKMQRALYELEIEGVQTNADFQLDLISDRNVIAGDYDTSFLMETFLPK
                    YQEKE"
gene            6322..7188
                    /gene="SP0426"
CDS             6322..7188
                    /gene="SP0426"
                    /note="identified by match to TIGR    protein family HMM
                    TIGR00515"
                    /codon_start=1
                    /transl_table=11
                    /product="acetyl-CoA carboxylase, carboxyl transferase,
                    beta subunit"
                    /protein_id="AAK74589.1"
                    /db_xref="GI:14971896"
                    /translation="MALFSKKDKYIRINPNRSVREKFQAKPEVPDELFSQCGCKHTI
                    YCKDLGSSRICPHCSYTFRISAQERLALTIDMGTFKELFTGISKDPLHFPGYQKKLA
                    SMREKTGLHEAVVTGTALIKGQTVALGIMDSNFIMASMGTVVGEKITRLFEYATVEKL
                    PVVLFTASGGARMQEGIMSLMQMAKISAAVKRHSNAGLFYLTILDPTTGGVTASFAM
                    EGDIILAEPQSLVGFAGRVIENTVRESLPEDFQKAEFLLEHGFVDAIVKRRDLPDTi
                    ASLVRLHGGSPR"
gene            7185..7952
                    /gene="SP0427"
CDS             7185..7952
                    /gene="SP0427"
                    /note="identified by match to TIGR    protein family HMM
                    TIGR00513"
                    /codon_start=1
                    /transl_table=11
                    /product="acetyl-CoA carboxylase, carboxyl transferase,
                    alpha subunit"
                    /protein_id="AAK74590.1"
                    /db_xref="GI:14971897"
                    /translation="MNIAKIVREAREQSRLTTLDFATGIFDEFIQLHGDRSPRDDGAV
                    VGGIGWLGDQAVTVVGIOKGKSLQDNLKRNFGQPHPEGYRKALRLMKQAEKFGRPVVT
                    FINTAGAYPGVGAEERGQGEAIARNLMEMSDLKVPIAIIGEGGSGALALAVADRV
                    WMLENSIYAILSPEGFASILWKDGTRAMEAAELMKITSHELLMDVVDKVISIGLSS
                    KELIKSVKKELQTELARLSQKFLEELLEERYQRFRKY"
gene            7963..8070
                    /gene="SP0428"
CDS             7963..8070
                    /gene="SP0428"
                    /note="identified by Glimmer2; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein"
                    /protein_id="AAK74591.1"
                    /db_xref="GI:14971898"
                    /translation="MRLERSHFLWKNVIQVEKLPVEEMGYKIDKKRNIL"
```

```
Alignment Scores:
Pred. No.:                              Length:             12039
                                        Matches:            306
Score:            3.83e-108
                  1518.00               Conservative:       0
Percent Similarity:      100.00%        Mismatches:         0
Best Local Similarity:   100.00%        Indels:             0
Query Match:             100.00%
```

VERSION   BD003687.1 GI:18631648
KEYWORDS   JP 2001501833-A/7.
SOURCE   unidentified
  ORGANISM   unidentified
             unclassified.
REFERENCE   1 (bases 1 to 19702)
  AUTHORS   Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
             Fannon,M. and Dougherty,B.A.
  TITLE   Polynucleotide of Streptococcus pneumoniae and sequence
  JOURNAL   Patent: JP 2001501833-A 7 13-FEB-2001;
             HUMAN GENOME SCIENCES INC
COMMENT   OS   Unidentified
             PN   JP 2001501833-A/7
             PD   13-FEB-2001
             PF   30-OCT-1997 JP 1998520718
             PR   31-OCT-1996 US   60/029960
             PI   CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI
             STEVEN C BARASH,
             PI   MICHAEL FANNON, BRIAN A DOUGHERTY
             PC   C12N15/09,A01K67/027,C07K16/12,C12N1/15,C07K16/12,C12N1/19,
             PC   C12N1/21,
             PC   C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
             G06F15/40
             CC   Strandedness: Double;
             CC   Topology: Linear;
             FH   Key   Location/Qualifiers
             FT   source   1..19702
             FT   /organism='Unidentified'.

FEATURES   Location/Qualifiers
     source   1..19702
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT   5736 a   4548 c   3728 g   5690 t
ORIGIN

Alignment Scores:
Pred. No.:   7.17e-108   Length: 19702
Score: 1518.00   Matches: 306
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: Gaps: 6

US-09-308-397-2 (1-306) x BD003687 (1-19702)

```
Qy      1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db  17149 ATGACTAAAACAGCCTTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG 17090

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db  17089 GATTTCTATGATCAGTATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGCCAGTGCTC 17030

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluLeuAspProLysLeuAsnGlnThrArgTyr 60
Db  17029 GGTTATGATTTGCGTTATCTCATCGATACGGAAGAGACAAACTCAATCAGACCCGCTAT 16970

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGluLysGly 80
Db  16969 ACGCAACCAGCCATCTCAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC 16910

Qy     81 TyrGlnProAspMetValAlaGlyGlyTyrSerAlaLeuValAlaSer 100
Db  16909 TATCAGCCTGATATGGTTGCTGGTGGTTACTCTGCCTTGGTGGCAAGT 16850

Qy    101 GlyAlaLeuAspPheGluAspProAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db  16849 GGCGCCTTGGATTTTGAAGATCCGGTTGCCTTGGTAGCTAAGCGTGGAGCCTATATGGAA 16790

Qy    121 GluAlaAlaProAlaAspSerGlyLysLysMetValAlaLeuAsnThrProValGluVal 140
Db  16789 GAAGCGGCTCCTGCGACTTCTGGCAAGAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC 16730
```

US-09-308-397-2 (1-306) x AE007354 (1-12039)

```
Qy      1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
Db   1057 ATGACTAAAACAGCCTTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG 1116

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db   1117 GATTTCTATGATCAGTAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGCCAGTGCTC 1176

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluLeuAspProLysLeuAsnGlnThrArgTyr 60
Db   1177 GGTTATGATTTGCGTTATCTCATCGATACGGAAGAGACAAACTCAATCAGACCCGCTAT 1236

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGluLysGly 80
Db   1237 ACGCAACCAGCCATCTCAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC 1296

Qy     81 TyrGlnProAspMetValAlaGlyGlyTyrSerAlaLeuValAlaSer 100
Db   1297 TATCAGCCTGATATGGTTGCTGGTGGTTACTCTGCCTTGGTGGCAAGT 1356

Qy    101 GlyAlaLeuAspPheGluAspProAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
Db   1357 GGCGCCTTGGATTTTGAAGATCCGGTTGCCTTGGTAGCTAAGCGTGGAGCCTATATGGAA 1416

Qy    121 GluAlaAlaProAlaAspSerGlyLysLysMetValAlaLeuAsnThrProValGluVal 140
Db   1417 GAAGCGGCTCCTGCGACTTCTGGCAAGAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC 1476

Qy    141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
Db   1477 ATTGAAGAAGCCTGTCAAAAAGCTTCTGAACTTGGAGTGGTTACTCCAGCCAACTATAAC 1536

Qy    161 ThrProAlaGlnIleValIleLeuAlaGlyLeuValValAlaValAspArgArgAlaValGluLeu 180
Db   1537 ACACCTGCACAAATCGTCATTGCTGGAGAGTGGTTGCAGTTGATCGAGCGGTTGAACTT 1596

Qy    181 LeuGluLysGluAlaGlyAlaLysLysArgLeuProLeuLysSerValSerGlyValProPheHisThr 200
Db   1597 TTGCAAGAAGCAGGTGCCAAACGCTTGATCCCTCTAAGGTGTCAGGTCCCTTTCACACC 1656

Qy    201 AlaLeuGluGluProAlaSerGlnLysProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
Db   1657 GCTCTCCCTGAACCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTTCA 1716

Qy    221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
Db   1717 GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGCTGTGATGCAAAAAGAGGACATT 1776

Qy    241 AlaGlnLeuLeuThrArgGlnValValLysProTyrGluSerIleGluVal 260
Db   1777 GCTCAGCTCTTGACGCGTCAGGTCGAAGGAACCCGTTCGTTCTATGAAAGTATTGGGGTC 1836

Qy    261 MetGlnGluGluIleSerArgAsnPheIleGluIleGlyProGlyLysProLysValLeuSerGly 280
Db   1837 ATGCAAGAAGAGGCATAAGCAACAACTTATCGAGATTGGACCCGGGGAAAGTTCTTGCAGGT 1896

Qy    281 PheValLysIleAspSerGlnThrAlaHisGlyLeuAlaHisValGluAspGlnAlaSerLeu 300
Db   1897 TTTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGGAGTTTA 1956

Qy    301 ValAlaLeuLeuGluLys 306
Db   1957 GTAGCACTTTTAGAAAAA 1974
```

RESULT 4
BD003687/c
LOCUS   BD003687   19702 bp   DNA   linear   PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION   BD003687

```
                                                        /serotype="19F"
                                                      . /db_xref="taxon:1313"
                                                        /clone="G54"

BASE COUNT    32875 a    20454 c    25072 g    32680 t    54 others
ORIGIN

Alignment Scores:
Pred. No.:            6.51e-107           Length:              111135
Score:                1518.00             Matches:             306
Percent Similarity:   100.00%             Conservative:        0
Best Local Similarity: 100.00%            Mismatches:          0
Query Match:          100.00%             Indels:              0
DB:                                       Gaps:                2

US-09-308-397-2 (1-306) x SPNEU1906 (1-111135)

Qy      1 MetThrLysThrAlaPheLeuPheLeuAlaGlyAlaGlnTyrLeuGlyMetGlyArg   20
Db  26478 ATGACTAAAACAGCTTTTTATTTGTGTGGTGCCCAGTATCTAGGGATGGGACGG  26537

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu   40
Db  26538 GATTTCTATGATCAGTATCCGATTGTCAAAGAAACGATTGAGCGAGCTCAGGTGCTC  26597

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyr   60
Db  26598 GGTTATGATTTACGTTATCTCATCGATACAGGAAGAGACAAAACTCAATACAGAGTAT  26657

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly   80
Db  26658 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC  26717

Qy     81 TyrGlnProAspMetValAlaGlyValLeuSerLeuGlyValSerAlaLeuValAlaSer  100
Db  26718 TATCGCCTGGATATGGTTGCTGGTGTTCTCTCTTTGGGAGAATACTCTGCCTTGGTGCAAGC  26777

Qy    101 GlyAlaLeuAspPheGluAspAlaLeuAlaLeuLeuAlaLeuLysArgGlyAlaTyrMetGlu  120
Db  26778 GGGCAGCCTTGAACTTGAAGATGCGGTTGCCTTGGTAGCGTGGAGCCTATATGGAA  26837

Qy    121 GluAlaAlaProAlaAspSerArgGlyLysMetValAlaValLeuAsnThrProValGluVal  140
Db  26838 GAAGCGGCTCCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC  26897

Qy    141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn  160
Db  26898 ATTGAAGAAGCCTGTCAAAAAGCTTCAGAACTTGGAGTGGTTACTCCAGCCAACTATAAC  26957

Qy    161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu  180
Db  27018 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGACCGAGCGGTTGAACTT  27017

Qy    181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr  200
Db  27018 TTGCAAGAAGCAGGTGCCAAACGTTTAATTCCTCTTAAGGTGTCAGGTCCCTTCACACC  27077

Qy    201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGlyLeuAlaGluThrLeuGlnValSerPheSer  220
Db  27078 GCTCTCCTTGAACCTGCTAGCCAAGAAACTAGCTCAGGTAAGTTTTCA  27137

Qy    221 AspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIle  240
Db  27138 GATTTTACTTGTCCCCTAGTGGCAATACAGAAGCTGTGATGCAAAAAGAGGACATT  27197

Qy    241 AlaGlnLeuLeuThrArgGlnValValLysGluProValArgPheTyrGluSerIleGluGlyVal  260
Db  27198 GCTCAGCTCTTGACGCGTCAGGTCAAGGAACCCGTTCGTTTCTATGAAAGTATTGGGTC  27257

Qy    261 MetGlnGluAlaGlyIleLeuSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly  280
Db  27258 ATGCAAGAAGCAGGCATAAGCAACTTTATCGAGATTGGACCGGGGAAAGTCTTGTCAGGT  27317

Qy    281 PheValLysIleLysAspGlnThrArgHisLeuAlaHisLeuAlaSerLeuLeu  300
```

```
Qy    141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn  160
Db  16729 ATTGAAGAAGCCTGTCAAAAAGCTTCAGAACTTGGAGTGGTTACTCCAGCCAACTATAAC  16670

Qy    161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAlaValGluLeu  180
Db  16669 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGACCGAGCGGTTGAACTT  16610

Qy    181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr  200
Db  16609 TTGCAAGAAGCAGGTGCCAAACGTTTAATTCCTCTTAAGGTGTCAGGTCCCTTCACACC  16550

Qy    201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGlyLeuAlaGluThrLeuGlnValSerPheSer  220
Db  16549 GCTCTCCTTGAACCTGCTAGCCAAGAAACTAGCTCAGGTAAGTTTTCA  16490

Qy    221 AspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGluAspIle  240
Db  16489 GATTTTACTTGTCCCCTAGTGGCAATACAGAAGCTGTGATGCAAAAAGAGGACATT  16430

Qy    241 AlaGlnLeuLeuThrArgGlnValValLysGluProValArgPheTyrGluSerIleGluGlyVal  260
Db  16429 GCTCAGCTCTTGACGCGTCAGGTCAAGGAACCCGTTCGTTTCTATGAAAGTATTGGGTC  16370

Qy    261 MetGlnGluAlaGlyIleLeuSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly  280
Db  16369 ATGCAAGAAGCAGGCATAAGCAACTTTATCGAGATTGGACCGGGGAAAGTCTTGTCAGGT  16310

Qy    281 PheValLysIleLysAspGlnThrArgHisLeuAlaHisLeuAlaSerLeu  300
Db  16309 TTTGTTAAAAAATTGATCAACTGCTCACTTAGCTCAGTCAGCAAGATCAAGCCGAGTTTA  16250

Qy    301 ValAlaLeuLeuGluLys  306
Db  16249 GTAGCACTTTTAGAAAAA  16232
```

```
RESULT 5
SPNEU1906
LOCUS        SPNEU1906        111135 bp    DNA      linear    HTG  11-JUL-2001
DEFINITION   Streptococcus pneumoniae clone G54. *** SEQUENCING IN PROGRESS ***,
             in ordered pieces.
ACCESSION    AL449928
VERSION      AL449928.1  GI:11545153
KEYWORDS     HTG; HTGS_PHASE2.
SOURCE       Streptococcus pneumoniae.
ORGANISM     Streptococcus pneumoniae
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE    1  (bases 1 to 111135)
AUTHORS      Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
             Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
             Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
             and Garcia-Bustos,J.F.
TITLE        Annotated draft genomic sequence from a Streptococcus pneumoniae
             type 19F clinical isolate
JOURNAL      Microb. Drug Resist. 7 (2), 99-125 (2001)
MEDLINE      21335329
PUBMED       11442348
REFERENCE    2  (bases 1 to 111135)
AUTHORS      Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
             Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
             Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
             Garcia-Bustos,J.F.
TITLE        Direct Submission
JOURNAL      Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
             Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT      * NOTE: This is a 'working draft' sequence.
             * This sequence will be replaced
             * by the finished sequence as soon as it is available and
             * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..111135
                     /organism="Streptococcus pneumoniae"
```

```
Db  27318 TTTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 ValAlaLeuLeuGluLysLys 306
Db  27378 GTAGCACTTTTAGAAAAA 27395                                    27377
```

RESULT 6
AX194275
LOCUS   AX194275   945 bp   DNA   linear   PAT 15-AUG-2001
DEFINITION Sequence 256 from Patent WO0149721.
ACCESSION   AX194275
VERSION   AX194275.1 GI:15211757
KEYWORDS   .
SOURCE   Streptococcus pneumoniae.
ORGANISM   Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 945)
AUTHORS   Dougherty,T.J., Pucci,M.J., Dougherty,B.A., Davison,D.B.,
            Bruccoleri,R.E. and Thanassi,J.A.
TITLE   Novel bacterial genes and proteins that are essential for cell
            viability and their uses
JOURNAL   Patent: WO 0149721-A 256 12-JUL-2001;
            Bristol-Myers Squibb Co. (US)
FEATURES   Location/Qualifiers
source   1..945
            /organism="Streptococcus pneumoniae"
            /db_xref="taxon:1313"
BASE COUNT 254 a   205 c   232 g   254 t
ORIGIN

Alignment Scores:
Pred. No.:
Score: 2.55e-109   Length: 945
Percent Similarity: 1515.00   Matches: 305
Best Local Similarity: 99.67%   Conservative: 1
Query Match: 99.80%   Mismatches: 0
DB: Indels: 0
                                                   Gaps: 0

US-09-308-397-2 (1-306) x AX194275 (1-945)

```
Qy    1 MetThrLysThrAlaPheLeuPheAlaGlyValGlnGlyAlaGlnTyrLeuGlyTyrMetGlyArg 20
Db    1 ATGACTAAAACAGCCTTTTTATTTGCTGGTGTCCAAGGTGCCCAGTATCTAGGGTATATGGGACGG 60

Qy   21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db   61 GATTTCATGATCAGTATCCCCATTGTTAAAGAAACGATTGATCGAGCGAGTCAGGTGCTA 120

Qy   41 GlyTyrAspArgTyrArgIleLeuAspThrGluAspLysLeuAsnGlnThrArgTyr 60
Db  121 GGTTATGATTTGCGTTATCTCATCGATACGGAAGATAAGCAAACTCAATCAGACCCGCTAT 180

Qy   61 ThrGlnProAlaIleLeuAlaLeuAlaThrSerValAlaIleTyrArgGlyLeuGlnGluLysGlyGly 80
Db  181 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC 240

Qy   81 TyrGlnProAspMetValAlaGlyValGlyLeuSerLeuGlyTyrSerAlaLeuValAlaSer 100
Db  241 TATCAGCCTGATATGGTTGCTGGTGTTGGTCTTCTTGGAGAATACTCTGCCTTGGTGGCAAGC 300

Qy  101 GlyAlaLeuAspPheGluAspPheAlaValAlaLeuLeuValAlaLysAlaArgGlyAlaTyrMetGlu 120
Db  301 GGCGCCCTTGAATTTGAAGATGCGGTTGCCCTGGTAGCCTTAAGCGGTGGAGCCTATATGGAA 360

Qy  121 GluAlaLeuAlaProAlaAspSerSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db  361 GAAGCCGGTCCTCCTGGACCTTCTGGCAAGATGGTAGCAGTGTTAAATACGCCAGTAGAGGTC 420

Qy  141 IleGluGluAlaCysGlnLysAlaSerGlyLeuLeuGlyValValThrProAlaAsnTyrAsn 160
Db  421 ATTGAAGAAGCCTGTCAAAAAGCTTCTGAACTTGGAGTGGTTACTCCAGCCAACTATAAC 480

Qy  161 ThrProAlaGlnIleValIleAlaGlyValValAlaAlaValAspAlaValAspAlaValGluLeu 180
Db  481 ACACCTGCACAAATCGTCATTGCTGGAGTGGTTGCAGTTGCTGACGCGGTTGACTT 540

Qy  181 LeuGlnGluAlaGlyValAlaLysAlaArgLeuIleProLeuValSerGlyValProPheHisThr 200
Db  541 TTGCAAGAAGCAGGTGTCGCCAAAGCTCGTTTGATTCCTCTTAAGGTGTCAGGTGCCCTTCACACC 600

Qy  201 AlaLeuLeuGluProAlaAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerProPheSer 220
Db  601 TCTCTCCTTGAACCTGCTAGCCAGAAACTAGCTGAAACTCTGGCTCAGGTAAGTTTTTCA 660

Qy  221 AspPheThrCysProLeuValGlyAsnThrGluGlyAlaAlaValMetGlnLysGluAspIle 240
Db  661 GATTTCACTTGTCCCCTAGTCGGCAATACAGAGGCTGCTGTGATGCAAAAAGAGGACATT 720

Qy  241 AlaGlnLeuLeuThrArgGlnLysLysGluProValAlaArgPheTyrGluSerIleGlyVal 260
Db  721 GCTCAGCTTCTTGACGCGGTCAGGTCAAGGAACCCGTTCGTTCTATGAAAGTATTGGGGTC 780

Qy  261 MetGlnGluAlaGlyLysIleSerAsnPheIleGlyIleLeuGlyProGlyLysValLeuSerGly 280
Db  781 ATGCAAGAAGCAGGCATAAGCAACTTTATCGAGATTGACCGGGGAAGTCTTGTCAGGT 840

Qy  281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
Db  841 TTTGTTAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTA 900

Qy  301 ValAlaLeuLeuGluLysLys 306
Db  901 GTAGCACTTTTAGAAAAA 918
```

RESULT 7
AE008418
LOCUS   AE008418   10925 bp   DNA   linear   BCT 13-SEP-2001
DEFINITION Streptococcus pneumoniae R6 section 34 of 184 of the complete
            genome.
ACCESSION   AE008418 AE007317
VERSION   AE008418.1 GI:15457935
KEYWORDS   .
SOURCE   Streptococcus pneumoniae R6.
ORGANISM   Streptococcus pneumoniae R6
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 10925)
AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
TITLE   Genome of the bacterium Streptococcus pneumoniae strain R6
JOURNAL   J. Bacteriol. 183 (19), 5709-5717 (2001)
MEDLINE   21423245
PUBMED   11544234
REFERENCE   2 (bases 1 to 10925)
AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
TITLE   Direct Submission
JOURNAL   Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
            Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES   Location/Qualifiers

```
source         1..10925
               /organism="Streptococcus pneumoniae R6"
               /strain="R6"
               /db_xref="taxon:171101"
repeat_region  complement(122..233)
               /rpt_family="BOX element"
               /rpt_type=dispersed
               /rpt_unit=Box_CBA
gene           264..1049
               /gene="phaB"
               /note="spr0375"
CDS            264..1049
               /gene="phaB"
               /EC_number="4.2.1.17"
               /note="Putative"
               /codon_start=1
               /transl_table=11
               /label=spr0375
               /product="Enoyl-CoA hydratase II"
               /protein_id="AAK99179.1"
               /db_xref="GI:15457936"
               /translation="MEHIIYQLEEDLALITLNRPEVANGFHIPMCEEILEALTLABEN
               PAVHFILINANGKVPSVGGDLVEMKRAVDEDDIPSLITKIAELVNTISYKIKQIAKPVL
               MEVDGAVAGAAANMAVAADFCLATDKAKFIQAFVGVGLAPDAGGIHLLSRSIGVTRAA
               QLAMTGEALTAEKALEWGLVYRVSEAEKLEKTREQLLKLRRASSNSYAAIKLVWES
               QFKDWQGYATLELNLQKSLAQTEDPFKEGVRAHSERRRPKFIGK"
gene           1143..1577
               /gene="spr0376"
CDS            1143..1577
               /gene="spr0376"
               /note="Probable transcriptional regulator"
               /codon_start=1
               /transl_table=11
               /label=spr0376
               /product="Conserved hypothetical protein"
               /protein_id="AAK99180.1"
               /db_xref="GI:15457937"
               /translation="MDYQRINEYLTSIFNNVLVIEEVNLRGSRFKDISIKEMTIDVI
               GKAPEDVTPSQVSKELMVTLGTVTTISLNNLERKGYIERVRSEQDRRVVHLHLTKKGRLI
               HRLHKRFHKAMVEKIIDGMSEEEIAVMGKGLTNLYQFLEDLK"
gene           1577..2551
               /gene="fabH"
               /note="spr0377"
CDS            1577..2551
               /gene="fabH"
               /EC_number="2.3.1.41"
               /codon_start=1
               /transl_table=11
               /label=spr0377
               /product="3-oxoacyl-acyl-carrier-protein synthase III"
               /protein_id="AAK99181.1"
               /db_xref="GI:15457938"
               /translation="MAFAKISQVAHYVPEQVVTNHDLAQIMDTNDEWISSRTGIRQRH
               ISRTSTSDCATEVAKKLMAKAGITGEELDFIILATITPDSMMPSTAARVQANIGANK
               AFAFDLTAACSGFVFALSTAEKFIASGRFQKGLVIGSETLSKAVDWSDRSTAVLFGDG
               AGGVLLEASQEHFLAESLNSDGSRSECLTYGHSGLHSPFSDQESADSFLKMDGRTVF
               DFAIRDVAKSIKQTIDESPIEVTDLDYLLLHQANDRILDKMARKIGVDRAKLPANMME
               YGNTSAASIPILLSECVEQGLIPLDGSQTVLLSGFGGGLTWGTLILTI"
gene           2611..2835
               /gene="acp"
               /note="spr0378"
CDS            2611..2835
               /gene="acp"
               /codon_start=1
               /transl_table=11
               /label=spr0378
               /product="Acyl carrier protein"
               /protein_id="AAK99182.1"
               /db_xref="GI:15457939"
               /translation="MAVFEKVQEIIVEELGKDASEVTLESTFDDLDADSLDLFQVISE
               IEDAPDIQIEAENDLKTVGDIVAYYEEQAK"
gene           2954..3928
               /gene="fabK"

CDS            /note="spr0379"
               2954..3928
               /gene="fabK"
               /codon_start=1
               /transl_table=11
               /label=spr0379
               /product="Enoyl-acyl carrier protein(ACP) reductase."
               /protein_id="AAK99183.1"
               /db_xref="GI:15457940"
               /translation="MKTRITELLKIDYPIFQQGMAWVADGDLAGAVSKAGGLGIIGGG
               NAPKEVVKANIDKIKSLIDKPFGVNIMLLSPFVEDIVDLVIEEGVKVVTTGAGNPSKY
               MERFHEAGIIVLPVVPSVALAKRMEKIGADAVIAEGMEAGGHIGKLTTMTLVRQVATA
               VSIPVIAGGIADGEGAAAGFMLGAEAVQGTRFVVAKESNAHPNYKEKILKARDIDT
               TISAQHFGHAVRAIKNQLTRDFELAEKDAFKQEDPDLEIFEOMGAGALAKAVVHGDVE
               GGSVMAGQIAGLVSKEETAEEILKDLYGAAKKIQEEASRWTGVVRND"
gene           3921..4841
               /gene="fabD"
               /note="spr0380"
CDS            3921..4841
               /gene="fabD"
               /EC_number="2.3.1.39"
               /codon_start=1
               /transl_table=11
               /label=spr0380
               /product="Malonyl acyl carrier protein transacylase"
               /protein_id="AAK99184.1"
               /db_xref="GI:15457941"
               /translation="MTKTAFLFAGQGACYLGMGRDFYDQYPIVKETIDRASQVLGYDL
               RYLIDTBEDKLNQTRYTQPAILATSVAIYRLLQEKGYQPDMVAGLSLGEYSALVASGA
               LDFEDAVALVAKRGAYMEEAAPADSGKMVAVLNTPVEVIEEACQKASBLGVVTPANYN
               TPAQIVIAGEVVAVDRAVELLQEAGAKRLIPLKVSGPHTSLLEPASQKLAETLAQVS
               ESDFTCPLVGNTEAAVMQKEDIAQLLTRQVKBPVRFYESIGVMQEAGISNFIEIGPGK
               VLSGFVKKIDQTAHLAHVEDQASLVALLEK"
gene           4875..5606
               /gene="fabG"
               /note="spr0381"
CDS            4875..5606
               /gene="fabG"
               /EC_number="1.1.1.100"
               /codon_start=1
               /transl_table=11
               /label=spr0381
               /product="3-ketoacyl-acyl carrier protein reductase"
               /protein_id="AAK99185.1"
               /db_xref="GI:15457942"
               /translation="MKLBHKNIFITGSSRGIGLAIAHKFAQAGANIVLNSRGAISEEL
               LAEFSNYGIKVVPISGDVSDFADAKRMIDQAIAELGSVDVLVNNAGITQDTIMLKMTE
               ADFEKVLKVNLTGAFNMTQSVLKPMMKARAGAIINMSSVGLMNSGIQCANYAARSAAGL
               IGFTKSVAREVASRNIRVNVIAPGMIESDMTAILSDKIKEATLAQIPMKEFGQAEQVA
               DUTVFLAGQDYLTGQVIAIDGGLSM"
gene           5619..6863
               /gene="fabF"
               /note="spr0382"
CDS            5619..6863
               /gene="fabF"
               /EC_number="2.3.1.41"
               /codon_start=1
               /transl_table=11
               /label=spr0382
               /product="Beta ketoacyl-acyl carrier protein synthase"
               /protein_id="AAK99186.1"
               /db_xref="GI:15457943"
               /translation="MKRMKLNRVVVTGVGVTSPIGNTPEEFWNSLATGKIGIGGITKF
               DHSDFDVHNAAEIQDFPFDKYFVKKDTNRFDNYSLYALYAAQEAVNHANLDVEALNRD
               RFGVIVASGIGGIKEIEDQVLRLHEKGPRKVKPMTLPKALPNMASGNVAMRFGANGVC
               KSINTACSSSNDAIGDAFRSIKFGFQDVMLVGGTEASITPAIAGFQALTALSTTEDP
               TRASIPFDKDRNGFVMGEGSGMVLVLESLEHAEKRGATILAEVVGYGNTCDAYHMTSPH
               PESQGAIKAIKLALEEAISPEQVAYYNAHGTSTPANEKGESGAIVALVGKEVPVSST
               KSFTGHLLGAAGVEAIVTIEAMRHNFVPMTAGTSEVSDYIEANVVGQGLEKEIPYA
               ISNTFGFGGHNAVLAFKRWENR"
gene           6866..7351
               /gene="accB"
               /note="spr0383"
```

CDS   6866..7351
       /gene="accB"
       /EC_number="6.4.1.2"
       /codon_start=1
       /transl_table=11
       /label=spr0383
       /product="Biotin carboxyl carrier protein of acetyl-CoA carboxylase"
       /protein_id="AAK99187.1"
       /db_xref="GI:15457944"
       /translation="MNLNDIKDLMTQFDQSSLREFSYKNGTDELQFSKNEARPVPEVA
       TQVAPAPVLATPSPVAPSAPAETVAEEVPAPEAEASVATEGNLVESPLVGVVYLAAGP
       DKPAFVIVGDSVKKGQTIVIIEAMKVMNEIPAPKDGVTEIIVSNEMVEFGKGLVRI
       K"

gene   7348..7770
       /gene="fabZ"

Alignment Scores:
Pred. No.:
Score: 5.79e-108   Length: 10925
Percent Similarity: 1515.00   Matches: 305
Best Local Similarity: 100.00%   Conservative: 1
Query Match: 99.67%   Mismatches: 0
DB: 99.80%   Indels: 0
                     1   Gaps: 0

US-09-308-397-2 (1-306) x AE008418 (1-10925)

Qy   1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaAlaGlnTyrLeuGlyMetGlyArg 20
Db 3921 ATGACTAAAACAGCCTTTTAATTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG 3980

Qy   21 AspPheTyrAspGlnTyrProLeuValLysGluThrIleAspArgAlaSerGlnValLeu 40
Db 3981 GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCCAGGTGCTA 4040

Qy   41 GlyTyrArgSerLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyr 60
Db 4041 GGTTATGATTGCGTTATCTCATCGATACGGAGAAGACAAACTCAATCAGACCCGCTAT 4100

Qy   61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGluLysGlyGly 80
Db 4101 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC 4160

Qy   81 TyrGlnProAspMetValAlaGlyLyrLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
Db 4161 TATCAGCCTGATATGGTTGCTGGTTTGTCTCTTGGAGAATACTCTGCCTTGGTGGCAAGC 4220

Qy 101 GlyAlaLeuAspPheGluLysAlaLeuAlaLeuAlaLysArgGlyAlaThrTyrMetGlu 120
Db 4221 GGCGCCTTGGATTTTGAAGATGCGCTTGCCTTGGCTAAGCGTGGAGCCTATATGGAA 4280

Qy 121 GluAlaAlaProLeuAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
Db 4281 GAAGGCGCTCTGCGACTCTGGCAAGATGTAGCAGTTCTCAATACGCCAGTAGAGGTC 4340

Qy 141 IleGluGluValAlaCysGlnLysGlyAlaSerGluLeuGlyValThrProAlaAsnTyrAsn 160
Db 4341 ATTGAAGAAGCCTGTCAAAAGTTCTGAACTTGGAGTGGTTACTCCAGCCAACTATAAC 4400

Qy 161 ThrProAlaGlnIleValIleLeuAlaGlyGluValAlaValAlaAspArgAlaValGluLeu 180
Db 4401 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTCGATCGAGCGGTTGAACTT 4460

Qy 181 LeuGlnAsnGlyAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyTyrProPheHisThr 200
Db 4461 TTGCAAGAACAGGGTGCCAAACGCTTGATTCCTCTTAAAGTGTCAGGTTCCCTTTCACACC 4520

Qy 201 AlaLeuLeuGluProAlaSerGlnLysGlnLysLeuAlaLeuThrLeuAlaGlnValSerPheSer 220
Db 4521 TCTCTTCCTTGAACCTGCTAGCCAGAAACTAGCTGAAACTTCTGGCTCAGGTAAGTTTTTCA 4580

Qy 221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGlyGluAspIle 240

Db 4581 GATTTTACTTGTCCCCTAGTCGGCAATACAGAGAAGTGTCTGTGATGCAAAAAGAGGACATT 4640

Qy 241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
Db 4641 GCTCAGGTCTTGACGCGTCAGGTCAAGGAACCCGTTCGTTTCTTCATGAAAGATATTGGGGTC 4700

Qy 261 MetGlnValAlaGlyIleGlySerAsnPheIleGluIleGlyProGlyLysValLysSerGly 280
Db 4701 ATGCAAGAAGCAGGCATAAGCAACTTTATCGAGATTGGACCGGGAAAGTCTTGTCAGGT 4760

Qy 281 PheValLysIleAspGlnThrAlaHisLeuAlaHisValGlnAspGlnAlaSerLeu 300
Db 4761 TTTGTTAAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGGAGTTTA 4820

Qy 301 ValAlaLeuLeuGluLys 306
Db 4821 GTAGCACTTTTAGAAAAA 4838

RESULT 8
AE006311
LOCUS   AE006311   10701 bp   DNA   linear   BCT 14-MAY-2001
DEFINITION   Lactococcus lactis subsp. lactis IL1403 section 73 of 218 of the complete genome.
ACCESSION   AE006311 AE005176
VERSION   AE006311.1 GI:12723690
KEYWORDS   .
SOURCE   Lactococcus lactis subsp. lactis.
ORGANISM   Lactococcus lactis subsp. lactis
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
REFERENCE   1 (bases 1 to 10701)
AUTHORS   Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K., Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE   The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403
JOURNAL   Genome Res. 11 (5), 731-753 (2001)
MEDLINE   21235186
PUBMED   11337471
REFERENCE   2 (bases 1 to 10701)
AUTHORS   Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarme,K., Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE   Direct Submission
JOURNAL   Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de Vilvert, Jouy en Josas 78352, France
FEATURES   Location/Qualifiers
source   1..10701
             /organism="Lactococcus lactis subsp. lactis"
             /strain="IL1403"
             /sub_species="lactis"
             /db_xref="taxon:1360"
             114..335
             /codon_start=1
gene   /gene="acpA"
             /transl_table=11
             /note="L0183"
             /product="acyl carrier protein"
CDS   114..335
             /protein_id="AAK04870.1"
             /gene="acpA"
             /db_xref="GI:12723691"
             /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
             /translation="MAVFEKVQDIIVDELGKEKEEVTLETSFEELDADSLDLFQIIND
             PHOSPHOLIPID METABOLISM. General. KEY COMPONENT IN DE NOVO
             IEDERDVEVDTEADMKTVADLVKYVENNK"
             FATTY ACID BIOSYNTHESIS"
gene   615..1541
             /gene="fabD"
             /note="L0184"
CDS   615..1541
             /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
             /gene="fabD"
             PHOSPHOLIPID METABOLISM. General. FATTY ACID BIOSYNTHESIS"
             /codon_start=1

```
          /transl_table=11
          /product="malonyl CoA-acyl carrier protein transacylase
          (EC 2.3.1.39)"
          /protein_id="AAK04871.1"
          /db_xref="GI:12723692"
          /translation="MTKTAFLFSGGQAQKLGMARDLYDQYETVKATFDEASQALGYDL
          RALIDNDEEKLNETKYTQPAILTTSVAILRLLSENGIKPDLVAGLSLGEYSALVASGI
          IDFQZAVKLVAKRGQYMTEAAPTGSGKMVAVMNTDPALIEEICQKAAEFKGGIVSPAN
          YNTPAQIVIGGEVEAVDYAVELLKEAGVRKLIELKVSGPFHTALIKPASEKLALELDK
          IDFRPPELPLISNTSAKVMENDEVKGLLTRQVMEPVRPVRFYESVETMQKLGATRFIEVGP
          GRVLSGFIKKIDKNAEIANVENLASFEALINQ"
gene      1627..2358
          /gene="fabG1"
          /note="L0185"
CDS       1627..2358
          /gene="fabG1"
          /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
          PHOSPHOLIPID METABOLISM. General. FIRST REDUCTION STEP IN
          THE FATTY ACID BIOSYNTHESIS PATHWAY"
          /codon_start=1
          /transl_table=11
          /product="3-oxoacyl-acyl carrier protein reductase (EC
          1.1.1.100)"
          /protein_id="AAK04872.1"
          /db_xref="GI:12723693"
          /translation="MEIKNKNFVFVTGSTRGIGKAIALQFAKAGSNLIINGRSAISEEL
          LAEFTAYGVKAVGISGDISKSEDAKQMVAEAIETLGSVDILVNNAGITRDGLLKMSE
          EDFBSVLKINLTGAFNMTQAVLKPMTRARSGAIINISSVVGLMGNAGQANYAASKAGL
          IGLTKSIAREVAARNVRVNAVAPGFIESDMTEVLSDKVKDAMKGQIPMKRFGMEEIA
          TATQPLAEQEYMTGQVLTIDGGVSM"
gene      2399..3628
          /gene="fabF"
          /note="L0186"
CDS       2399..3628
          /gene="fabF"
          /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
          PHOSPHOLIPID METABOLISM. General. FATTY ACID BIOSYNTHESIS"
          /codon_start=1
          /transl_table=11
          /product="3-oxoacyl-acyl carrier protein synthase II (EC
          2.3.1.41)"
          /protein_id="AAK04873.1"
          /db_xref="GI:12723694"
          /translation="MTNRVVISGYGVVSAIGNSPEEFWNNLKAGKTGIGPITHFDAEA
          TGISVAAEVXEFPFDKYFQKKDARRMDTFSLYAVVAALDAMEMSGITEENTNFDRLGC
          IMASGIGGLEQSQKNSQAIVAKGPRKGVAPLYVPLAIANMATGNVALRTGARGVSRAE
          VTACAAGANSIGSAFREIKHGYADAMIAGGAEAAICELGIAGFANLTALITKETDPEKA
          CIPFDKNRSGFVMGEGSGVLILESLEHAQARGANILAEIVGYGNTNDAFHMTTPSGVG
          AENAIKLALEAGASPADVDYVNAHGTSHANEETESKAIHNVIGDKAVVSSTKALHG
          HALGATGAIEAVATVQALLNQYAPVNAGTTELDEGMTINVVLGEGKPADINLALSQDF
          GFGGHNAVLAFKKWTGE"
gene      3635..4102
          /gene="accB"
          /note="L0187"
CDS       3635..4102
          /gene="accB"
          /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
          PHOSPHOLIPID METABOLISM. General. FIRST STEP IN LONG-CHAIN
          FATTY ACID SYNTHESIS"
          /codon_start=1
          /transl_table=11
          /product="biotin carboxyl carrier protein of acetyl-CoA
          carboxylase (EC 6.4.1.2)"
          /protein_id="AAK04874.1"
          /db_xref="GI:12723695"
          /translation="MNISEVKDLMNQPNGSTLREFSWKRADGELSFSKNEGQALVGAQ
          SAAPTARPAFTAQQATPVABSTEAPVASEAAVEAEGEAVTSPLVGVAYLKPSPDKAEFA
          KVGGDSVKKGQTLLIEAMKVMNEIPAPKDGVITEIMVSGEDVIEFGQDLMRIK"
gene      4229..4663
          /gene="fabZ2"
          /note="L0188"
CDS       4229..4663
          /gene="fabZ2"

          /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
          PHOSPHOLIPID METABOLISM. General. FATTY ACID BIOSYNTHESIS"
          /codon_start=1
          /transl_table=11
          /product="3R-hydroxymyristoyl-(3R)-HYDROXYMYRISTOYL-acyl
          carrier protein dehydratase (EC 4.2.1.1)"
          /protein_id="AAK04875.1"
          /db_xref="GI:12723696"
          /translation="MTEVNINVTEIMEALPHRYPFLLVDRVIDIAEDEITAIKNVTIN
          EEFFQGHFPQYPVMPGVLIMEALAQAAGVLELSKPENKGKLVFYAGMDNVKYKKQVTP
          GDQLVLHAKFIKRRGPIAVVEAEATVDGKLAAKGTLTFALGR"
gene      5057..6424
          /gene="accC"
          /note="L0189"
CDS       5057..6424
          /gene="accC"
          /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
          PHOSPHOLIPID METABOLISM. General. FIRST STEP IN LONG-CHAIN
          FATTY ACID SYNTHESIS"
          /codon_start=1
          /transl_table=11
          /product="biotin carboxylase (EC 6.3.4.14)"
          /protein_id="AAK04876.1"
          /db_xref="GI:12723697"
          /translation="MFNKILIANRGEIAVRIIRAARELGVGTVAIYSTADRDALHVHL
          ADEAICVGPARSTRSYLMNNILSAAVATGAQAIHPGFGFLSENSKFARLCEEMNIKF
          IGPSAKVMDMGDKINARAEMIKAGVPVTPGSEGEVHTDEAVKIAEEIGYPIMLKAS
          AGGGGKGIRKVNNVEELVPAFEAAQAEAQAAFGNGAMFIERMIFPARHIEVQLGDSQ
          GNVIHLGERDCSLQRNQKVLEESPSVAIGHTLRNEITAAAVKAAQHVSYENAGTIEF
          LLDESSGKFYFMEMNTRVQVEHPVTEFVSGVDIVKEQIRVAGNELSVKQEDVTFSGH
          AIECRINAENPKKNFAPSPGKINLFLPSGGVGLRVDSAMYSGYSIPPYYDSMIAKVI
          VHGNRFEALMKQRALLEFDVGGVITNVDFQLELISDPHVVAGDYDTSFLGNVFLPE
          YQKD"
gene      6510..7376
          /gene="accD"
          /note="L0180"
CDS       6510..7376
          /gene="accD"
          /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
          PHOSPHOLIPID METABOLISM. General. FATTY ACID BIOSYNTHESIS"
          /codon_start=1
          /transl_table=11
          /product="acetyl-CoA carboxylase carboxyl transferase
          subunit betta (EC 6.4.1.2)"
          /protein_id="AAK04877.1"
          /db_xref="GI:12723698"
          /translation="MALFQKKKYIKINPNRSIIEKQAEQPEVDELPAKCPACKHTIY
          QKDLGKNKVCPNCDYNFRTAKERLAIVADKGSFVEMFTGIESKNPLDFPGYPEKLAA
          TKARTGLDEAVITGATKGQKTALAIMDSTFMASMGTVVGEKLRLFEYATTEKLP
          IIVFTASGGARWQEGIMSLMQMAKTSAAVKRHSNAGLFYITVLTDPTGGVTASFASL
          GDIILAEBQSLIGPAGRRVIEQTVRQTLPDDFQKAEFLLNHGFVDAIVKRTELRQKLA
          LLLELHTEVENV"
gene      7369..8151
          /gene="accA"
          /note="L0190"
CDS       7369..8151
          /gene="accA"
          /note="EVIDENCE BY HOMOLOGY BIO07.01 FATTY ACID AND
          PHOSPHOLIPID METABOLISM. General. FIRST STEP IN LONG-CHAIN
          FATTY ACID SYNTHESIS"
          /codon_start=1
          /transl_table=11
          /product="acetyl-CoA carboxylase carboxyl transferase
          subunit alpha (EC 6.4.1.2)"
          /protein_id="AAK04878.1"
          /db_xref="GI:12723699"
          /translation="MSETIDSLTGEVSQKNIADIINERARAADRISARDIIDKVFTDFF
          ELRGDRQLTDDEAIVGIARPNNQAVTVIGIQKGRNLAENLATNFGQPSPNGYRKALR
          LMKQAEKFGRPVTTFINTAGAYPGIEAERGQGEAIARNLLEMSDLKVPIIAFITGEG
          GSGGALALAANKVYILENAVYSVLSPEGFATIILWKDGSRRDEAAELMKITAPHLLEM
          GLVDAIISBENLKRQLTSLGDLSLLTETELAEQRYERFRKY"
gene      8280..9422
          /gene="metB2"
```

```
Db      1515 |||  ||||||||||||||:::  ::::  :::
             TCATTGAAGCTTTGATTAATCAG 1538
```

```
                              /note="L0181"

US-09-308-397-2 (1-306) x AE006311 (1-10701)

Alignment Scores:
Pred. No.:
Score:                   1.18e-71
                         1048.00
Percent Similarity:      80.52%
Best Local Similarity:   67.53%
Query Match:             69.04%
DB:

Length:         10701
Matches:        208
Conservative:   40
Mismatches:     58
Indels:         2
Gaps:           1
```

```
Qy      1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
          |||:::||||:||||||::  |::| ::|   ||| ||:|||::||| ||::| || |
Db      615 ATGACTAAAACAGCATTTTTATTCTCAGGTCAAGGGCACAAAAGCTTGGAATGGCACGT 674

Qy     21 AspPheTyrAspGlnTyrProIleValLysGlyUThrIleAspArgAlaSerGlnValLeu 40
          |||:||||  ||| |||||  || :||| :|||::|||| |||||:||| | |  ::|||
Db      675 GACTTATATGACCAATATGAAACAGTTAAAGCAACTTTGATGAAGCAAGTCAAGCTTTA 734

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluLysLeuAsnGlnInThrArgTyr 60
          |:: |||::|||||||  ||::||||  ||| :|::|||| :|| |||| | ::|||
Db      735 GGATATGCATTTGCGAGCTTGATTGATAATGATGAAGAAAAACTAAATGAAACAAAGTAC 794

Qy     61 ThrGlnProAlaIleLeuUAlaThrSerValAlaIleTyrArgLeuLeuGlnLeuLysGlyGly 80
          |:: ||  |:||:||:|||| :|||| |||   | |||| ::| |:| ::||| :|||:|
Db      795 ACTCAACCTGCAATTTTAACAACCTCGTGCTATTTACGGTTGTTAAGTGGAGAATGGG 854

Qy     81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaAlaSer 100
          : :||   ::: | :|| |||:||| ||  ||||||||| : |||||: |  |||||||
Db      855 ATTAAACCTGACCTTGGTGCGTCCTGTACTTGGTGAATATTCTGCCTTGGTAGCATCA 914

Qy    101 GlyAlaLeuAspPheGluGlyAspAlaValAlaGlyAlaGlyAlaTyrMetGlu 120
          |::| :|:| ||:| |||||| |||:|:|| :||  | ||| | |  |||:||:
Db      915 GGAATCATTGATTTTCAAGAAGCAGCTAAGCTTGCAGCTAAACGTGGTCAATATGACA 974

Qy    121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGluVal 140
          |:: |:| | |:|  | |||:: |||| :|::|||||:|| ::|:||:| ::||||
Db      975 GAAGCTGCACCGACTGGTTCGTCGGTAAAATGTCGCTGTAATGACACCAGACCCAGCCTTG 1034

Qy    141 IleGluGlnAlaCysGlnLysAlaSerGluLeu-----GlyValValThrProAlaAsn 158
          |:: |||:| | ||: | || ||||||  ||||     ||:|||| ||:|||:||||::
Db     1035 ATTGAGCAAAATTGCCAGAAAGCGGCGGCCGAATTAAAGGCGTATTGTGAGGTCCAGCAAAT 1094

Qy    159 TyrAsnThrProAlaGlnIleIleValIleAlaGlyGluValValAlaValAlaAspArgAlaVal 178
          : :|||:| |  ||:|:|| :|||:|| | |:| |||||||:|| :||::|||||::
Db     1095 TATAAACGCCCGCACAAATTGTTATTGGTGGTGAAGTTGGAGCCGGTTGATTATGCTGTT 1154

Qy    179 GluLeuLeuGlyGluGlnGluProLeuProLeuLysProLeuGlySerGlyProPhe 198
          |:: :|||:  |::|  |||:|| | ||::|| ||| |::| :| |:|:| || |||
Db     1155 GAGTTGCTAAAAGAAGAGCGGAGTTCGTAAACTTATTGAATTAAAAGTTTCAGGAGACCTTTC 1214

Qy    199 HisThrAlaLeuLeuGluProAlaSerGlnLysLeuLeuLysAlaGlyThrGluAlaGlnInValSer 218
          | |:|  | ::|:|| ||  |||::|:| |||| |||:|| :|| |||::|::|:|:|:
Db     1215 CATACAGCAATTTTAAAACCAGCATCTGAAAAATTGGCTTTGAAGGCTGGTAAAAATTGAT 1274

Qy    219 PheSerAspPheThrCysProLeuValGlyAlaAsnThrGlnAlaAlaValMetGlnLysGlu 238
          ||:|  | |:: ||||   ||:|:||| |:| |||:|:| | |  ::| |||:|
Db     1275 TTTAGACCTTTGAATTACCATTAATCTCAAATACCAGTGCTAAAGTAATGAAAATGAT 1334

Qy    239 AspIleAlaGlnInLeuLeuThrArgGlnValLysGluTyrGluSerIle 258
          ||: |:| ||| |::| :|:| :|| ||::| |:: || ::|:::|:::|
Db     1335 GAAGTCAAGGACTTTTGACCGGTCAAGTCATGGAACCAGTTCGTTTTTATGAAATCGGTT 1394

Qy    259 GlyValMetGlnGlnAlaGlyIleSerAsnPheIleGlnIleGlyProGlyLysValLeu 278
          |::|:||   :|| | || ||:|   ||:|:|: :||:|| |||:|:|||::|||:|
Db     1395 GAAACAATGCAAAAACTAGGGCGACTCGCTTTATTGAAGTTGGTCCTGGGAGTACTT 1454

Qy    279 SerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAla 298
          |::| :||||::|::|  ||:|:|:| |:: || ||:  ::| :|| ||:|:|||::||
Db     1455 TCAGGTTCATTAAAAAATTGATAAAAATGCAGAAATTGCTAACTTTGAAAATTTAGCT 1514

Qy    299 SerLeuValAlaLeuLeuGluGluLys 306
```

gene

/translation="MALFRKKDKYIRITPNNSLKGSVSHNVPEVPDELFAKCPACKHM
IYKKDLGLAKICPTCSYNFRISAQERLTLTVDEGSFQELFTSISTKDPLRFPGYQEKL
QKAKETTGLHEAVLTGKAMVKEQKIALAIMDSHFIMASMGTTVVGEKITRLFELAIEEN
LPVVIFTASGGARMQEGIMSLMQMAKVSAAVKHSNAGLFYLTILTDPTTGGVTASFA
MEGDIILAEPQSLVGFAGRRVIETTVRENLPDDFQKAEFLQDHGFVDAIVKRTELRDK
IAHLVAFHGGGQ"
complement(1784. .3148)
/gene="accC"
/note="SPy1745"
complement(1784. .3148)

CDS

/gene="accC"
/note="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel
electrophoresis from another strain of S. pyogenes. Hogan,
D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen.
2000. Two-dimensional gel electrophoresis map of
Streptococcus pyogenes proteins. Unpublished data.; Best
Blastp hit = pir||A69581 acetyl-CoA carboxylase (biotin
carboxylase subunit) accC - Bacillus subtilis
>gi|1303913|dbj|BAA12569.1| (D84432) YqhX [Bacillus
subtilis] >gi|2634868|emb|CAB14365.1| (Z99116) acetyl-CoA
carboxylase subunit (biotin carboxylase subunit) [Bacillus
subtilis]"
/codon_start=1
/transl_table=11
/product="putative acetyl-CoA carboxylase biotin
carboxylase subunit"
/protein_id="AAK34449.1"
/db_xref="GI:13622800"
/translation="MFKKLIIANRGEIAVRIIRAARELGISTVAVYSEADKEALHTIL
ADEAICIGPARSKESYLNMNSVLSAAIVTGAQAIHPGFGFLSENSKFATMCEENNIKF
IGPSASVMDKMGDKINAREEMIKAGVPVIPGSDEVYNAQEALAIANKIGYPVMLKAS
AGGGKGIRKVETEADLEAAFNAASQEALGAFGNGAMYLEKVIYPARHIEVQIIGDAY
GNIIHLGERDCSLQRNNQKVLEESPSIAIGNTLRIEHGQAAVRAEAVAYENAGTIEF
LLDEDSEKFYFMEMNTRIQVEHPVTEFVTGVDIVKEQIKIAAGQPLAINQEDITITGH
AIECRINAENTAFNFAPSPGKITDLYMPSGGVGLRVDSAVYNGYAIPPYYDSMIAKII
VHGSNRFDAIMKMQRALVELEIEGIITNTDFQLDLISDKRVIAGYDTSFLMETFLPH
YNKD"
complement(3180. .3599)
/gene="fabZ"
/note="SPy1746"
complement(3180. .3599)

gene

/gene="fabZ"
/note="Best Blastp hit = dbj|BAB16042.1| (AB030883)
similar to Bacillus subtilis hydroxymyristoyl- (acyl
carrier protein) dehydratase [Streptococcus
zooepidemicus]"
/codon_start=1
/transl_table=11
/product="putative beta-hydroxyacyl-ACP dehydratase"
/protein_id="AAK34490.1"
/db_xref="GI:13622801"
/translation="MDIREIQAALPHRYPMLLVDRVLEVSDDHIVAIKNVTINEPFFN
GHFPHYPVMPGVLIMEALAQTAGVLELSKEENKGKLVFYAGMDKVKFKKQVPGDQLV
MTATFKRRGTIAVVEARAERVDGKLAASGTLTFACGQ"
complement(3599. .4099)
/gene="accB"
/note="SPy1747"
complement(3599. .4099)

gene

/gene="accB"
/note="Best Blastp hit = pir||C75558 acetyl-CoA
carboxylase, biotin carboxyl carrier protein - Deinococcus
radiodurans (strain R1)
>gi|6457783|gb|AAF09708.1|AE001874.5 (AE001874) acetyl-CoA
carboxylase, biotin carboxyl carrier protein [Deinococcus
radiodurans]"
/codon_start=1
/transl_table=11
/product="putative biotoin carboxyl carrier protein"
/protein_id="AAK34491.1"
/db_xref="GI:13622802"
/translation="MNIQEIKDLMAQFDTSSLREFLFKTNEGELIFSKNEQHLNASTS
NQEHAVPVPVQVLVPNPTAGSEASSPASVKDVPVEEQPQAESFVAEGDIVESPLVGVAY

gene

LAASPDKRPPFVAVGDTVKKGQTLVIIEAMKVMNEVPAPCDGVITEILVSNEDVIEFGQ
GLVRIK"
complement(4101. .5333)
/gene="fabF"
/note="SPy1748"
complement(4101. .5333)

CDS

/gene="fabF"
/EC_number="2.3.1.41"
/note="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel
electrophoresis from another strain of S. pyogenes. Hogan,
D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen.
2000. Two-dimensional gel electrophoresis map of
Streptococcus pyogenes proteins. Unpublished data.; Best
Blastp hit = pir||G69842 3-oxoacyl-[acyl-carrier-protein]
synthase (EC 2.3.1.41) - Bacillus subtilis
>gi|2633471|emb|CAB12975.1| (Z99109) similar to 3-oxoacyl-
acyl-carrier protein synthase [Bacillus subtilis]
>gi|2633488|emb|CAB12991.1| (Z99110) similar to 3-oxoacyl-
acyl-carrier protein synthase [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative beta-ketoacyl-ACP synthase II"
/protein_id="AAK34492.1"
/db_xref="GI:13622803"
/translation="MTFKRVVVTGYGLTSPIGHDPETFWNNLKAGQIGIGPITKFDTT
DYAVKNAAEIQDFPPDKYFVKKDLNRFDRYSIYALYAAKEAINHADLNIEMVDSDRFG
VIVASGIGIAEIEEQVIRLHEKGPRKVKPMTLPKALPNMAAGNVAMSLKAQGVCKSI
NTACASSNDAIGDAPRAIKFGTQDVMIVGGSEAAITKFAIAGFQSLTALSTEDPSRS
SIPDKDRNGFIMGEGSGMLVLESLEHAQERGATILARIVGYGRTAILAEIVDCAYHMTSNPEG
LGARKAIHLALQEAGIEASAINYVNAHGTSTPANEKGKSGQAIVAVLGKKVPVSSTKSF
TGHLLGAAGAIEATATEAMRHNYVPMTAGTQALSEDIEANVIFGEGKETAINYAISN
TFGFGGHNAVLAFKCWEE"
complement(5348. .6082)
/gene="fabG"
/note="SPy1749"
complement(5348. .6082)

gene

/gene="fabG"
/EC_number="1.1.1.100"
/note="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel
electrophoresis from another strain of S. pyogenes. Hogan,
D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen.
2000. Two-dimensional gel electrophoresis map of
Streptococcus pyogenes proteins. Unpublished data.; Best
Blastp hit = dbj|BAB06210.1| (AP001515)
3-oxoacyl-(acyl-carrier protein) reductase [Bacillus
halodurans]"
/codon_start=1
/transl_table=11
/product="putative beta-ketoacyl-ACP reductase"
/protein_id="AAK34493.1"
/db_xref="GI:13622804"
/translation="MEIKGKNIFITGSTRGIGLAMAHQPASLEANIVLNGRSAISEEL
VASFTDYGVTVVTISGDVSEASEAKRMVNEAIESLGSIDVLVNNAGITNDKLMLKMTE
EDERVLKINLTGAFNMTQSVLKPMIKARQGAINVSSVVGLTGNIGQANYAASKAGM
IGFTKSVAREVAARNICVNAIAPGFIESDMTGVLPEKMQEQILSQIPMKRIGKAQEVA
HLASFLVEQDYITGQVIAIDGGMTMQ"
complement(6072. .7010)

gene

Alignment Scores:		
Pred. No.:	Length:	10909
Score:	1.04e-70	Matches:	209
Percent Similarity:	1036.00		
Best Local Similarity:	81.05%	Conservative:	39
Query Match:	68.30%	Mismatches:	58
DB:	68.25%	Indels:	0
	1	Gaps:	0

US-09-308-397-2 (1-306) x AE006603 (1-10909)

Qy   1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg   20
                                                                        |||
Db 7010 ATGACAAAGACAGCCTTTTTATTTGCCGGTCAAGGTGCTCAAAATTAGGATGGCAAGG 6951

```
REFERENCE   1 (bases 1 to 50416)
AUTHORS     Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
            Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
            Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
            Musser,J.M.
TITLE       Genome sequence of a serotype M3 strain of group A Streptococcus:
            Phage-encoded toxins, the high-virulence phenotype, and clone
            emergence
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
REFERENCE   2 (bases 1 to 50416)
AUTHORS     Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
            Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
            Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
            Musser,J.M.
TITLE       Direct Submission
JOURNAL     Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
            Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
            Hamilton, MT 59840, USA
FEATURES             Location/Qualifiers
     source          1..50416
                     /organism="Streptococcus pyogenes MGAS315"
                     /strain="MGAS315"
                     /serotype="M3"
                     /db_xref="taxon:198466"
                     /note="group: A"
     gene            117..1103
                     /gene="SpyM3_1493"
     CDS             117..1103
                     /gene="SpyM3_1493"
                     /note="best blastp hit: gb|AAL98256.1| (AE010083) putative
                     esterase [Streptococcus pyogenes MGAS8232]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative esterase"
                     /protein_id="AAM80100.1"
                     /db_xref="GI:21905237"
                     /translation="MKHPIRKTLVTLGLLLTLCLPTPAAASSRSWKSWFIEQYFWLKR
                     DKSYYSKQDDPSFQRYLDACREQSDKPYQLDTNLVNGPLVQENLYGMQVYSWNDNGKP
                     DQKTIYLAGGSYLNNPTTYHINMLKTLSTSLDAKIVLPIYPKAPRYTYNYTMPKLVN
                     LYQHYHKNQNVFLMGDSAGGGLALGLAHALHNESVPQPKQLVLLSPWLDVTMSHPEI
                     PKYEDADPILSSWGLKRVGELMAYSADNTNHIYVSPKNGPITYLPPITLFGTREIFY
                     PDIRDYAAKLKAANHNITFITQEGMNHVYPIYPIEEAKTAQYQIIDAINKTP"
     gene            complement(1239..1589)
                     /gene="rbfA"
                     /note="SpyM3_1494"
     CDS             complement(1239..1589)
                     /gene="rbfA"
                     /note="best blastp hit: gb|AAK34465.1| (AE006601) putative
                     ribosome binding factor A [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative ribosome binding factor A"
                     /protein_id="AAM80101.1"
                     /db_xref="GI:21905239"
                     /translation="MANHRIDRVGMEIKREVNDILQKKVRDPRVQGVTITEVQMQGDL
                     SLAKVYTIMSDLASDNQKAQTGLEKATGTTKRELGKQLTMYKIPDLVFEKDNSIAYG
                     NKIDQLLRDLDNKS"
     gene            complement(1795..4599)
                     /gene="infB"
                     /note="SpyM3_1495"
     CDS             complement(1795..4599)
                     /gene="infB"
                     /note="best blastp hit: gb|AAK34466.1| (AE006601) putative
                     initiation factor 2 [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative initiation factor 2"
                     /protein_id="AAM80102.1"
                     /db_xref="GI:21905240"
                     /translation="MVEHAKYLGLDVKSHASSVEEADAKKIISSFSKASKPDVTASQT
                     VKPKEVAQPSVTVVKETGSEHVEKTQVSKPKSRNFKAERRARAKEQAARKQANGSSHR
                     SQERRGGYRQPNNHCTNEQGDKRITHRSQGDTNDKRIERKASNVSPRHDNHQLVGDRN
                     RSFAKENHKNGRFTNQKKQGEPQQKSPKIDFKARAAALKAEQNAEYSRQSETRFRA
```

```
Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
             |||||||:: ::   ||||||||  |:||||| ||:|:|   |||:|||||| |||
Db   6950 GATTTTTATGATAACTTGCCATTGTGAAGAAAAACCTTGATCAAGCTAGTCAAGTATTG 6891

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGlnAsnLysLeuAsnGlnThrArgTyr 60
             |||| |||  |||  ::|||||  |||:||| ||: | :|||||| |||| |:| |
Db   6890 GGATACGGATTGCGCCGTTTGATTGATACGGAGCAGAGTAAAACTTAATCAAACTAGCTAT 6831

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnLysGlyGly 80
             ||   |||   :|:  |||    ||:|||:||  |||||| |||  ||||| | |||
Db   6830 ACGCACCAGCTATTTGACATCATCAATTGCTATTTACCGTGTTTTGGGCTTACATCAC 6771

Qy     81 TyrGlnProAspMetValAlaGlyValLeuSerLeuGlyTyrSerAlaLeuValAlaSer 100
             |||||||  |:|||| ||| || ||   |:|    || | :|||| |||: || :|
Db   6770 GTTAAACCGGATATGGTAGCTGGTCTTCCTTAGGAGAATACTCAGCTTGGTAGCATCA 6711

Qy    101 GlyAlaLeuAspPheGluAspAlaAlaLeuValAlaLeuLysArgGlyAlaTyrMetGlu 120
              ||:||  |:||||||:||| ||:| ||||| ||:||| |||::|||| | :||||||
Db   6710 GGGGCACTCTCTTTTGAAGATACCCTATCCTTAGTAGCTAAGAGGCCGCTTGATGGAG 6651

Qy    121 GluLeuAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
              ||:||| ||  |  |   ||| ::|: :||:|||||:||||||:|| ||||| |||
Db   6650 GAAGCAGCACCACAAGGATCCGGAAAGATGTTGCCGTTATGAATACAGACGGTGCAAGTC 6591

Qy    141 IleGluGluAlaCysGlnLysGlySerGluLeuGlyValValAlaThrProAlaAsnTyrAsn 160
             |||||||| | ||: |||| | ||:|||:|||||:|| ||   || |||| ||  ||
Db   6590 ATCGAAGAAGCTCGTCAAATAGCTCGAAAATGGTGTTGCTAAAACATACTATAAT 6531

Qy    161 ThrProAlaGlyIleLeuAlaGlyIleGluValAlaValAlaValAlaAspArgAlaValGluLeu 180
             || ||| | ||    ||| |||:| :|| |: ||  |||||||| || ||| |||  ||
Db   6530 ACTCCTAGTCAAATTGTAATTGGTGGTCAGACAGATGCTGGTGCAGTGAACCTT 6471

Qy    181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyValProPheHisThr 200
             ||: |||||||  |   ||| |:  | |||  |   |: |  ||  ||||  | :||||
Db   6470 TTAAAGGAAGCAGGGGAGTTAAGCGTTTAATCCCTTTAAACGTGTCAGGTCTTTCCACACT 6411

Qy    201 AlaLeuGluProAlaSerGlnLysValAlaLeuGluThrLeuValAlaGluAlaGlnValSerPheSer 220
             |||:||| | ||  |:|||| ||| ||:||:|| ||| |: |:    :|||  ||
Db   6410 GCTTTGTTAGAACCAGCTAGCCGCTTGTTGGCTAAGAGTTGAAAGATACAACTTCAGT 6351

Qy    221 AspPheThrCysProLeuValGlyAsnThrGlyAlaAlaValMetGlnLysGluAspaspIle 240
             ||| |  ||| ||:  |||  |||||||  |||  |:|||| |||   |||| ||
Db   6350 GACTTCAAGATTCCTTTGGTTGGTAATACCGAAGCTAATATTATGGAAAAAGACCGTATC 6291

Qy    241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
             ||:|||:|  :|  |   | |||:| ||  |:|||  ||:|: ||  |: |||||||
Db   6290 CCAGAACTATTAGCCCGTCAAGTCATGAGCCTGTTCGTTTTTATGACAGTGTTGCGACT 6231

Qy    261 MetGlnGluAlaGlyIleSerAsnPheIleGluLeuGlyProGlyLysValLeuSerGly 280
             :: |||| | |:|   |:|:  |:|:  ||| | || | |||||  || ||||| |
Db   6230 TTAGTAGAAGTGGCATAACACCAATTCATTGAGGTAGGACCAGGTAAGGTTTTGACAGGT 6171

Qy    281 PheValLysIleAspGlnThrArgAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
             |||||||| ||  |:||:||:|||:| |: |||:|  ||  |||||  | |: |||:||
Db   6170 TTTGTTGAAGAAAATTGATAAAAATTTACTACTAGTGTTGAAAACATGGTAGCTTA 6111

Qy    301 ValAlaLeuLeuGluLys 306
             |:  |||||||:
Db   6110 CGTTTATTTTTAGATAGA 6093
```

```
RESULT 10
LOCUS       AE014165/c      50416 bp    DNA     linear   BCT 19-JUL-2002
DEFINITION  Streptococcus pyogenes MGAS315, section 30 of 37 of the complete
            genome.
ACCESSION   AE014165 AE014074
VERSION     AE014165.1  GI:21905237
KEYWORDS    .
SOURCE      Streptococcus pyogenes MGAS315
  ORGANISM  Streptococcus pyogenes MGAS315
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
```

```
                    QQBAKRLABLARQEAKEAALKAQAEMSHREAALKSIEEAETKLKSSNISAKSTADNR
                    RKKQARPEKNRELTIHSQEGQKKNKKSWNSQNQVRNQKNSNWNKKTKKGKNAKNTN
                    TAPKPVTERKFHELPKEFEYTEGMTVAEIAKRIKREPAEIVKLFMMGVMATONQSLD
                    GDTIELLMVDYGIEAKAKVEVDDADIERFEDENYLNPENIVERAPVVTIMGHVDHGK
                    TTLLDTLRNSRVATGEAGGITGHIGAYQIEBAGKKITFLDTPGHAAFTSMRARGASVT
                    DITILIVAADDGVMPQTIEAINHSKAAGVPIIVAINKIDKPGANPERVIAEBLAEYGII
                    STAWGGECEFVEISAXFNKNIDELLETVLLVAEVBELKADPTVRAIGTVIEARLDKGK
                    GAIATILVQQGTLHVQDPIVVGNTFGRVRAMVNDLGRRVKSAEPSTPVSITGLNETPM
                    AGDHFAVYADEKAARAAGEERSKRALLKQRQNTQRVSLDNLFDTLKAGEIKTVNVIIK
                    ADVQGSVEALAASLVKIEVEGVRVNVHSAVGAINESDVTLAEASNAVIIGFNVRPTP
                    QARQQADTDDVEIRLHSIIYKVIEBVEEAMKGKLDPVYQEKILGEAIIRETFKVSKVG
                    TIGGPMVINGKVTRDSSVRVIRDSVVIFDGKLASLKHYKDDVKEVGNAQEGGLMIENF
                    NDLKVDDTIEAYIMEEIVRK"
gene                complement(4676..4978)
                    /gene="SpyM3_1496"
CDS                 complement(4676..4978)
                    /gene="SpyM3_1496"
                    /note="best blastp hit: gb|AAL99259.1| (AE010083) putative
                    ribosomal protein [Streptococcus pyogenes MGAS8232]"
                    /codon_start=1
                    /transl_table=11
                    /product="putative ribosomal protein"
                    /protein_id="AAM80103.1"
                    /db_xref="GI:21905241"
                    /translation="MTNLERASSLIGLAQRAGKVISGEELVVKAIQHQQVILVPLAND
                    AGPNVTKKVTDKSNYNVEVSTVLNALELSAALGKPRKVAAIADAGFSKXMRTLME"
gene                complement(4971..5267)
                    /gene="SpyM3_1497"
CDS                 complement(4971..5267)
                    /gene="SpyM3_1497"
                    /note="best non-GAS blastp hit: gb|AAK74710.1| (AE007365)
                    conserved hypothetical protein [Streptococcus pneumoniae
                    TIGR4], and gb|AAK99283.1| (AE008427) Conserved
                    hypothetical protein [Streptococcus pneumoniae R6]"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAM80104.1"
                    /db_xref="GI:21905241"
                    /translation="MSKVKKIPLRKSLVSGEIIAKRDLRIVKTKDGQVFIDPTGKQN
                    GRGAIYKLDNQEALMAKKKQVFNRSFSMDIPESFYDDLIAYVDHKIKRELGLD"
gene                complement(5283..6428)
                    /gene="nusA"
                    /note="SpyM3_1498"
CDS                 complement(5283..6428)
                    /gene="nusA"
                    /note="best blastp hit: gb|AAK34469.1| (AE006601)
                    transcription termination-antitermination factor
                    [Streptococcus pyogenes M1 GAS]"
                    /codon_start=1
                    /transl_table=11
                    /product="putative transcription
                    termination-antitermination factor"
                    /protein_id="AAM80105.1"
                    /db_xref="GI:21905243"
                    /translation="MLEAFRLEEEKHIDKADIIDAVTESLKSAYKRRYGQSESCVIE
                    FNEKTADFQVFTVREVVEEVFDSRLEISLKDALAISSAYELGDKIRFEBSVNEFGRVA
                    AQSAKQTIMEKMRRQMREVMFNEYKEHEGEIMTGTVERFDQRFIYVNLGSLEAQLSHQ
                    DQIPGETFKSHDRIDVYVYKVENNPKGVNVFVSRSHPEFIKRIMEQEIPEVFDGTVEI
                    MSVSREAGDRTKVAVRSHNPNVDAIGTIVGRGGSNIKKVISKFPHPKRVDAKTGLEIPV
                    EENIDVIQWVDDPAEFIYNAIAPAEVDMVLFDDEDLKRATVVVPDSKLSLAIGRRGQN
                    VRLAAHLTGYRIDIKSASEYDRLEAEKEAATAVEBPVVDDAMTABEV"
gene                complement(6615..7151)
                    /gene="SpyM3_1499"
CDS                 complement(6615..7151)
                    /gene="SpyM3_1499"
                    /note="best non-GAS blastp hit: gb|AAK99281.1| (AE008427)
                    Conserved hypothetical protein [Streptococcus pneumoniae
                    R6]"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAM80106.1"
```

```
                    /db_xref="GI:21905244"
                    /translation="MDSGPIILEKSIKIEBEVIKIANTSIIDIVTKVTVPEIKAPYEL
                    VDVSYDKMGSDYILSILVDKEDGITVEDTSDLTNIISPLLDTIDPDFPNQYMLEVSS
                    PGLERPLKTADSLKAAVGSYINVSLYQAIDKVKVFQGDLLAFPGETLTIDYLDKTRHK
                    IVNIPYQAVAKVRMAVKL"
tRNA                complement(7259..7345)
                    /product="tRNA-Ser"
gene                complement(7410..8045)
                    /gene="SpyM3_1500"
CDS                 complement(7410..8045)
                    /gene="SpyM3_1500"
                    /note="best blastp hit: gb|AAL99263.1| (AE010083) putative
                    methyltransferase [Streptococcus pyogenes MGAS8232]"
                    /codon_start=1
                    /transl_table=11
                    /product="putative methyltransferase"
                    /protein_id="AAM80107.1"
                    /db_xref="GI:21905245"
                    /translation="MRVRKRKGAEEHLANNPHYVILNPEDAKGRWHDVFGNDRPIHIE
                    VGSGKGGFITGWALKNPDINYIGIDIQLSVLSVALDKVLASEVPNVKLLRVDGSSLTN
                    YFEDGEVDMYLNFSDPWPKTHGEKKRLTYKDFLPDTYKRILPEHGEIHFKTDNRGLFE
                    YSLASFSQYGMTLRQIWLDLHASNYEGNVMTEYEKFSNKGQVIYRVEANF"
gene                complement(8045..8836)
                    /gene="SpyM3_1501"
CDS                 complement(8045..8836)
                    /gene="SpyM3_1501"
                    /note="best blastp hit: gb|AAK99279.1| (AE008427)
                    Conserved hypothetical protein (Streptococcus pneumoniae
                    R6)"
                    /codon_start=1
                    /transl_table=11
                    /product="conserved hypothetical protein"
                    /protein_id="AAM80108.1"
                    /db_xref="GI:21905246"
                    /translation="MTTTEBQELTLTPLRGKSGKAYKGTYPNGBCVFIKLNTTPILPAL
                    AKEOIAPQLLWAKRMGNGDMMSAQEWLNGRTLTKEDMNSKQIIHLLRLHKSKKLVNQ
                    LLQLNYKIENPYDLLVDFEQNAPLQIQQNSYLQAIVKELKRSLPEFKSEVATIVHGDI
                    KHSNWVITTSGMIFLVDWDSVRLTDRMYDVAYLLSHYIPRSRWSEWLSYYGYKNNDKV
                    MQKIIWYGQFSHLJQILKCFDKRDMEHVNQEIYALRKFREIFRKK"
gene                complement(8900..9934)
                    /gene="SpyM3_1502"
CDS                 complement(8900..9934)
                    /gene="SpyM3_1502"
                    /note="best blastp hit: gb|AAK34473.1| (AE006601) putative
```

Alignment Scores:
```
Pred. No.:              7.32e-70          Length:          50416
Score:                  1036.00           Matches:         209
Percent Similarity:     81.05%            Conservative:    39
Best Local Similarity:  68.30%            Mismatches:      58
Query Match:            68.25%            Indels:          0
DB:                                       Gaps:            0
```

US-09-308-397-2 (1-306) x AE014165 (1-50416)

```
Qy      1 MetThrLysThrAlaPheLeuAlaGlyValGlnTyrLeuGlyMetGlyArg    20
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30296 ATGACAAAGACAGCCTTTTATTTGCCGGTCAAGGTGCTCAAAAATTAGGATGGCAAGG 30237

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu   40
            ||||||| :::|||||||||||||   ||::|:||| ::||||
Db  30236 GATTTTTATGATAACTTGCTATTGGAAGAAAAACTTTGATCAAGTCAGTCAAGTATTG 30177

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr   60
            ||||||||||||||||   :::||:|::||||||||:|||
Db  30176 GGATACGGATTTGCGCCGTTGATTGATAGTGACAGAGTTAAACTTAATCAAACAGACTAT 30117

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGlnLysLysGly   80
            |||:|||:|:||||   :::|::|::||:|:||
Db  30116 ACGCAACCAGCTATTTGACATCATCAATTGCTATTACCGTGTTTGGGCTTACATCAC 30057

Qy     81 TyrGlnProAspMetValAlaGlyValGluTyrSerAlaLeuValAlaSer   100
            :::|||||||||||
```

```
Db   30056 GTTAAACCGGATATGTAGCTGGTCGTTCCTTAGGAGAATACTCAGCTTTGGTAGCATCA 29997
                  ||||||| ||||||| :::||||||||||||||||||||||||||||||||||||
Qy     101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLeuValAlaArgGlyAlaTyrMetGlu 120
                  |||||||||||||
Db   29996 GGGGCACTCCTTTGAAGAACCCTATCCTTAGTAGCCTAAGAGAGGGCCGCTGATGAGG 29937

Qy     121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
                  :::|||||||||||| |||| :::||
Db   29936 GAAGCAGCACCACAAGGATCTGGAAAGATGGTTGCCGTTATGAATAACAGACGTGCAAGTC 29877

Qy     141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
                  |||||||| ||:::||
Db   29876 ATCGGAAGAAGTCTGTCAAATAGCTGCTAAACATGGAGTTGCTCCAGCAAACTATAAT 29817

Qy     161 ThrProAlaGlnIleValIleAlaGlyValValAlaAspArgAlaValGluLeu 180
                  ||| :::|||||| ||||
Db   29816 ACTCCTAGTCAAATTGTAATTGGTGGTCAGACAGATGCTGCAGTTGAACTT 29757

Qy     181 LeuGlnGluAlaGlyLeuAlaLysArgLeuIleProLeuLysValSerGlyIleProPheHisThr 200
                  ||:::||
Db   29756 TTAAAGGAAGCGGGAGTTAAGCGTTAATCCCTTTAAACGTGTCAGGTTCCTTTCCACACT 29697

Qy     201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
                  |||||:::  ||||:::
Db   29696 GCTTTGTTAGAACCAGCTAGCCGCTTGGCTAAAGAGTTGGAAAGATACAACTTCAGT 29637

Qy     221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
                  ||||||||||:::|||||
Db   29636 GACTTCAAGATTCCTTTGGTGGTAATACCGAAGCTAATATTATGGAAAAAGACCGTATC 29577

Qy     241 AlaGlnIleLeuThrArgGlnInValLysGluProValArgPheTyrGluSerIleGlyValVal 260
                  :::||||||||
Db   29576 CCAGAACTATTTAGCCCGTCAAGTCAGGAGCCCTGTTCGTTTTATGACAGTGTTGCGACT 29517

Qy     261 MetGlnGluAlaGlyIleLeuSerAsnPheLeuGlyIleGlyProGlyLysValLeuSerGly 280
                  :::||||| ||||||:::|||
Db   29516 TTAGTAGAAGTGGCATAACACAATTCATTGAGGTAGGACCAGGTAAGGTTTTGACAGGT 29457

Qy     281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGlnAlaSerLeu 300
                  ||||||  :::
Db   29456 TTTGTGAAGAAAATTGATAAAAATTTACTATGTACTAGTGTTGAAAAACATGGTAAGCTTA 29397

Qy     301 ValAlaLeuLeuGluLys 306
                  |||:::::
Db   29396 CGTTTATTTTTTAGATAGA 29379
```

```
RESULT 11
AE010088/c
LOCUS       AE010088      14456 bp      DNA        linear    BCT 03-APR-2002
DEFINITION  Streptococcus pyogenes strain MGAS8232, section 136 of 173 of the
            complete genome.
ACCESSION   AE010088 AE009949
VERSION     AE010088.1  GI:19748917
KEYWORDS    .
SOURCE      Streptococcus pyogenes MGAS8232.
  ORGANISM  Streptococcus pyogenes MGAS8232
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 14456)
  AUTHORS   Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
            Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
            Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
            Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
  TITLE     Genome sequence and comparative microarray analysis of serotype M18
            group A Streptococcus strains associated with acute rheumatic fever
            outbreaks
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)
   MEDLINE  21927593
   PUBMED   11917108
REFERENCE   2  (bases 1 to 14456)
  AUTHORS   Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
            Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
            Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
```

```
  TITLE     Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.George. and Musser,J.M.
  JOURNAL   Direct Submission
            Submitted (31-JAN-2002) Laboratory of Human Bacterial
            Pathogenesis/Rocky Mountain Laboratories/NIAID/NIH, 903 S. 4th St.,
            Hamilton, MT 59840, USA
FEATURES             Location/Qualifiers
     source          1..14456
                     /organism="Streptococcus pyogenes MGAS8232"
                     /strain="MGAS8232"
                     /db_xref="taxon:186103"
     misc_feature    <1..101
                     /note="phi_SDA"
     misc_feature    86..101
                     /note="phi Sda attR"
     gene            105..917
                     /gene="spyM18_1809"
     CDS             105..917
                     /gene="spyM18_1809"
                     /note="best blastp match gb|AAK34481.1| (AE006602)
                     conserved hypothetical protein [Streptococcus pyogenes M1
                     GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAL98329.1"
                     /db_xref="GI:19748918"
                     /translation="MTKKLIAIDLDGTLLHHDNTISTYQKTIKAVQDKGHHVIISTG
                     RPYRMALDYIQLNLKTPIITFNGALTHMPEQKWAYEHNVTLDKGYLRLLLKYQDDFQ
                     MDFIASEYRKQNVYIMTNPESIDPQLFGVDEITQDMALEITKITRNPNALLMQTHHED
                     KYALAKNMRAFFKDEIEIDSWGGPNIELEISSKNVNKAYALNYLLGIYNMDKKDLIAF
                     GDEHNDTEMLAFAGTGYAMKNASPVLLPYADQQLNFSNEDGVAKKLELFL"
     gene            1269..2261
                     /gene="spyM18_1810"
     CDS             1269..2261
                     /gene="spyM18_1810"
                     /note="best blastp match gb|AAK34482.1| (AE006602)
                     mannose-specific phosphotransferase system component IIAB
                     [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="mannose-specific phosphotransferase system
                     component IIAB"
                     /protein_id="AAL98330.1"
                     /db_xref="GI:19748819"
                     /translation="MGIGIIIASHGKFAEGIHQGGSMIFGEQEKVQVVTFMPNEGPDD
                     LYGHFNNAIQQFDADDEILVLADLWSGSPFNQASRVAGENPDRKMAIITGLNLPMLIQ
                     AYTERLMDAGAGVEQVAANIIKESKDGIALPEDLNPVEETAATEKVVNALQGAIPAG
                     TVIGDGKLKNLARVDTRLLHGQVATAWTPASKADRIIVASDEVAQDDLRKQLIKQAA
                     PGGVKANVVPISKLIEASKDPRFGNTHALILFQTPQDALRAVEGGVEINELNVGSMAH
                     STGKTMVNNVLSMDKEDVATFEKIRDLGVTFDVRKVPNDSKKNLFELIQKANIK"
     gene            2350..3159
                     /gene="spyM18_1811"
     CDS             2350..3159
                     /gene="spyM18_1811"
                     /note="best blastp match gb|AAK34483.1| (AE006602)
                     putative mannose-specific phosphotransferase system
                     component IIC [Streptococcus pyogenes M1 GAS]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative mannose-specific phosphotransferase
                     system component II"
                     /protein_id="AAL98331.1"
                     /db_xref="GI:19748920"
                     /translation="MSDISIISAILVVIIAFFAGLEGILDQFQMHQPLVACTIGLVT
                     GHLEAGVILGGTLQMLALGWANIGAVARPDAALASVARAIIMVKSGDFTQKGIIFAYS
                     TAIPLAVAGLFLTMIVRTLSTALVHAGDKAAEGNFAGIERFHFIALLLQGLRIAVPA
                     ALLVAVPTSAVQSVLNAMPNWLNEGMQIGGAMVVAVGYAMVINNMATREVWPFFALGF
                     ALAAISQLTLIAMGVIGVAIAFIYLNLSKKGGNGNAAGSADPIGDILEDY"
     gene            3176..4087
                     /gene="spyM18_1812"
     CDS             3176..4087
                     /gene="spyM18_1812"
                     /note="best blastp match gb|AAK34484.1| (AE006602)
```

```
                              putative mannose-specific phosphotransferase system
                              component IID [Streptococcus pyogenes M1 GAS]"
                              /codon_start=1
                              /transl_table=11
                              /product="putative mannose-specific phosphotransferase
                              system component IID"
                              /protein_id="AAL98332.1"
                              /db_xref="GI:19748921"
                              /translation="MTEQIKLTKSDRQRVWWRSQFLQGSWNYBRMQNMGWAYALIPAL
                              KKLYTSPEDRAAALERHMEFFNTHPYVAAPIIGVTLALEEERANGTPIDDKAIQGVKI
                              GMMGPLAGIGDPVFWFTVRPILGALGASLASTGNIVGPLLFFFGWNLIRMAFLWYTQE
                              FGYKAGSEITKDMSGGILQDITKGASILGMFIIALVLVQRWVSINFTIDLPGKQLSDGA
                              YVVFPDGAVKGAELKTILANAIGGMSLDKVQAQTLQGQLDSLIPGLAGLLITFLCMWL
                              LKKKVYSPIAIIIGLLPAFGIIAHLAGIM"
   gene                       4201..4560
                              /gene="spyM18_1813"
   CDS                        4201..4560
                              /gene="spyM18_1813"
                              /note="best blastp match gb|AAK34485.1| (AE006602)
                              conserved hypothetical protein [Streptococcus pyogenes M1
                              GAS]"
                              /codon_start=1
                              /transl_table=11
                              /product="conserved hypothetical protein"
                              /protein_id="AAL98333.1"
                              /db_xref="GI:19748922"
                              /translation="MAQSLNTSVEYKTKAVSYLGMGGKVGHILLGDKALEFYNDKNVN
                              DYIQIPWTAINHIGANVSRKKVSRHFEIFTDQGKFLFASGDSGKILKITRQHIGNEKV
                              ITLPTLMQTFINKFPRH"
   gene                       4952..6229
                              /gene="serS"
                              /note="spyM18_1814"
   CDS                        4952..6229
                              /gene="serS"
                              /note="best blastp match gb|AAK34486.1| (AE006602)
                              putative seryl-tRNA synthetase [Streptococcus pyogenes M1
                              GAS]"
                              /codon_start=1
                              /transl_table=11
                              /product="putative seryl-tRNA synthetase"
                              /protein_id="AAL98334.1"
                              /db_xref="GI:19748923"
                              /translation="MLDLKIRTDFDTVAAKLKNRGVSEDTLTHLKELDEKRTLLVQ
                              SEELKAERNIASAATAQAKRQKEPATQQIADMOKVSADIKTIDNQLVAIDQOVADIIT
                              VLPNPHDSVPVGADEEDNVEIRRWGTPRDFDFEVKAHWDLGEDLDILDWERGAKVTG
                              ARFLFYKNLGARLERALYNFMLDEHIKEGYQEIITPYMVNHDSMFGTQYPKFEDTF
                              ELADTNFVLIPTAEVPLTNYYRGEILDGKELPYIFTAMSPFRSEAGSAGRDTRGLIR
                              LHQFHKVEMVKFAKPEESYQELEKWTANAENILQKLGLPYRVISLCTGDMGFSAAKTY
                              DLEVNIPAQNTYREISSCGSNTEDFRURRAQIRYRDEADGKVKLHTLNGSGLAVGRTV
                              AAILENYQNEDGSVTIPEVLRPYMGGETVISPK"
   gene                       complement(6462..7232)
                              /gene="accA"
                              /note="spyM18_1815"
   CDS                        complement(6462..7232)
                              /gene="accA"
                              /note="best blastp match gb|AAK34487.1| (AE006603)
                              putative acetyl-CoA carboxylase alpha subunit
                              [Streptococcus pyogenes M1 GAS]"
                              /codon_start=1
                              /transl_table=11
                              /product="putative acetyl-CoA carboxylase alpha subunit"
                              /protein_id="AAL98335.1"
                              /db_xref="GI:19748924"
                              /translation="MTDVSRVLKEARDQGRLTTLDYANLIFDDFMEBLHGDRHFSDDGA
                              IVGGLAYLAGQPVTVGIQXGKNLQHNLARNFGQPNEGYRKALRLMXQABKFGRPVV
                              TPINTAGAYPGVGAEERGQEAIAKNLMEMSDLKVPIIAIIIGEGGSGALALAVADQ
                              VWMLENTMYAVLSPEGPASILWKDGSRATEAAELMKITAGELYQMGIVDRIIPEHGYF
                              SSEIVDIIKANLIEQITSLQAKPLDQLLDERYQRFRKY"
   gene                       complement(7229..8095)
                              /gene="accD"
                              /note="spyM18_1816"
   CDS                        complement(7229..8095)
                              /gene="accD"
```

```
                              /note="best blastp match gb|AAK34488.1| (AE006603)
                              putative acetyl-CoA carboxylase beta subunit
                              [Streptococcus pyogenes M1 GAS]"
                              /codon_start=1
                              /transl_table=11
                              /product="putative acetyl-CoA carboxylase beta subunit"
                              /protein_id="AAL98336.1"
                              /db_xref="GI:19748925"
                              /translation="MALPRKKDKYIRITPNNFLKGSVSHNVPEVPDELFAKCPACKHM
                              IYKKDLGLAKICPTCSYNFRISAQERILTLTVDEGSFQBLFTSIETKDPLRFPGYQEKL
                              QKAKETTGLHEAVLTGKAMVKGQOIALAIMDSHFIMASMGTVVGEKIRTLFLELAIEEN
                              LPVIFTASGGARMQEGIMSLMQMAKVSAAVKRHSNAGLFYLTILTDPTTGGVTASFA
                              MEGDIILAEPQSLVGFAGRRVIETTVRENLPDDFQKAEFLQDHGFVDAIVKRTELRDK
                              IAHIVAFHGGGQ"
   gene                       complement(8104..9468)
                              /gene="accC"
                              /note="spyM18_1817"
   CDS                        complement(8104..9468)
                              /gene="accC"
                              /note="best blastp match gb|AAK34489.1| (AE006603)
                              putative acetyl-CoA carboxylase biotin carboxylase subunit
                              [Streptococcus pyogenes M1 GAS]"
                              /codon_start=1
                              /transl_table=11
                              /product="putative acetyl-CoA carboxylase biotin
                              carboxylase subunit"
                              /protein_id="AAL98337.1"
                              /db_xref="GI:19748926"
                              /translation="MFKKILIANRGEIAVRIIRAARELGISTVAVYSEADKEALHTIL
```

```
Alignment Scores:
Pred. No.:                              14456
Score:                   4.36e-70       Matches:             208
Percent Similarity:      103.00         Conservative:         39
Best Local Similarity:   80.72%         Mismatches:           59
Query Match:             67.97%         Indels:                0
DB:                      67.85%         Gaps:                  0
```

```
US-09-308-397-2 (1-306) x AE010088 (1-14456)

Qy      1 MetThrLysThrAlaPheLeuPheAlaGlyAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13330 ATGACAAAAACAGCTTTTTATTTGCCGGTCAAGGTGCCCAAAAATTAGGGATGCGCAAGG  13271

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
          :||||||| | |||||||||| ||| |||||||||| |||||||||||||||||||||||
Db  13270 GATTTTTATGATAACTTTGCTATTGTGAAGAAAAACTTTTGATCAAGCTAGTCAAGTATTG  13211

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspGlyLysLeuAsnGlnThrArgTyr  60
          ||||||||||||||||||||||||||||:|:||||||||||||||||||:||||||||:||
Db  13210 GGATACGATTTGCGCGTTTGATCGATACGGAAGAGGACGGTAAAACTTAATCAAACTAGCTAT  13151

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGluLysGly  80
          ||||||||||||||||||||||||:||||:|||||||||||||||||:|||||||:||
Db  13150 ACGCAACCAGCTATTTTGACATCAATTGCTATTTACCGTGTTTTGGGCTTACATCAC  13091

Qy     81 TyrProAspMetValAlaGlyMetGlyTyrLeuSerLeuGlyTyrSerAlaLeuValAlaSer  100
          |||||||||||| ||||||||||||||||| |||||||||||||||||||||||||||
Db  13090 GTTAAACCGGATATGGTAGCTGGGTCTTTCCT7AGGAGAGATACTTGGTAGCATCA  13031

Qy    101 GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLeuArgGlyAlaTyrMetGlu  120
          |||||||||||||||||||||:||||||:|||||||||||||||||||||||||||||||
Db  13030 GGGGCACTCTCTTTTGAAGATACCCTATCTTAGTAGCCAAGAGAGGCCGCTTGATGGAG  12971

Qy    121 GluAlaAlaProAlaAspSerGlyLysMetValAlaLeuAsnThrProValGluVal  140
          ||||||||||||||||||:||:||||||||||||||||||||||||||||||||||||||
Db  12970 GAAGCAGCACCAGAAGATCTGGAAAGATGGTTGGCCGTTATGAATACAGACGTGCAAGTC  12911

Qy    141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn  160
          |||||||||| |||||||||||||||| |||||||| |||||||||||||||||||:|
Db  12910 ATCGAAGAAGTCGTCAAATAGCTGCTAAACATGGAGTGGTTGCTCCAGCAAACTATAAT  12851

Qy    161 ThrProAlaGlnIleValIleAlaGlyGluValValAlaValAlaAspArgAlaValGluLeu  180
```

```
Db  12850 GCTCCAGTCAAATTGTAATTGGTGGTCAGACAGATGCTGTGAACGTGGCAGTTGAACTT 12791
              |||||  :::: ::|||
Qy    181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProHisThr  200
Db  12790 TTAAAGGAAAGGGGAGTTAAGCGTTAATCCCTTTAAACGTGTCAGGTCCTTCCACACT 12731
Qy    201 AlaLeuLeuGluProAlaSerGlnLysValGluThrLeuAlaGluThrLeuValSerPheSer  220
Db  12730 GCT:::TGTTAGAACCAGCTAGCCGCTTGTTGCTAAGAGTTGGAAAGATACAAACTTCAGT 12671
Qy    221 AspPheThrCysProLeuValGlyAsnThrGluAlaValMetGlnLysGlyGluAspIle  240
Db  12670 GACTTCAAGATTCCTTTGGTGTAATACCCGAAGCTAATATTATGGAAAAAGACCGTATC 12611
Qy    241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal  260
Db  12610 CCAGAACTATTAGCCCGTCAAGTCATGAGCCTGTTTCGTTTTATGACAGTGTTGCGACT 12551
Qy    261 MetGlnGluAlaSerGlnIleSerAsnPheIleGluIleGlyProGlyLysValSerGly  280
Db  12550 TTAGTAGAAAGTGGCATAACACAATTCATTGAGGTAGGACCAGGTAAGTTTTGACAGGT 12491
Qy    281 PheValLysIleAspGlnThrAlaHisLeuAlaHisAlaValGluAspGlnAlaSerLeu  300
Db  12490 TTTGTGAAGAGAAATTGATAAAAAATTTACTATGTACTAGTGTTGAAAACATGGTAAGCTTA 12431
Qy    301 ValAlaLeuLeuGluLys  306
Db  12430 CGTTTATTTTTAGATAGA 12413
```

Alignment Scores:
Pred. No.:
Score: 4.33e-58
                      858.00
Percent Similarity: 98.37%
Best Local Similarity: 97.83%
Query Match: 56.52%
DB:

Length: 1196
Matches: 180
Conservative: 1
Mismatches: 2
Indels: 6
Gaps: 0

US-09-308-397-2 (1-306) x AR193756 (1-1196)

```
Qy    125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla  144
              :::||
Db      2 GCTGAACTTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTCATTGAAGAAGCC   61
Qy    145 CysGlnLeuGlyAlaSer-GluLeuGlyValValAlaThrProAlaAsnTyrArgAlaGl  164
              :::|||
Db     62 TGTCAAAAGTTCGGAACTTGGGAACTGGGTAGTGGTTACTCCAGCCAACATAACACCTGCACA  121
Qy    164 nIleValIleAlaGlyGlyGluValLysValAlaAlaValAlaAspArgAlaValGlyLeuLeuGlnGluAl  184
Db    122 AATCGTCATTGCTGGAGGAGTGGTTGCAGTTGATCGAGCGGTTGAACTTTTGCAAGAAGC  181
```

```
Qy    184 aGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGl  204
              ||| :::|||
Db    182 AGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCAGGTCCGTTCCACACCGCTCTCCTTGA  241
Qy    204 uProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCy  224
              |||
Db    242 GCCTGCTAGCCAGAAACTAGCTGAAACCTTAGCTCAGGTAAGTTTTTCAGATTTTACTG  301
Qy    224 sProLeuValGlyAsnThrGluAlaValMetGlnLysGlyGluAspIleAlaGlnLeuLe  244
Db    302 TCCCCTAGTCGGCAATACAGAGAGCTGCGTGTGATGCAAAAAGGAGACATTGCTCAGCTCTT  361
Qy    244 uThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGlnGluAl  264
Db    362 GAGCGTCAGGTCAAGAAACCCTGGTTCTATGAAAGTATTGGGGTCATGCAAGAAGC  421
Qy    264 aGlyIleSerAsnPheIle-GluIleGlyProGlyLysValLeuSerGlyPheValLysL  284
Db    422 AGGCATAAGCAACTTTATTGAGATTGGACCCGGGAAAGTCTTGTCAGGTTTTGTTAAAA  481
Qy    284 ysIleAspGlnThrAlaHisLeuAlaHisAlaValGluAspGlnAlaSerLeuValAlaLeuL  304
Db    482 AAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTTT  541
Qy    304 euGluLys  306
Db    542 TAGAAAAA  549
```

```
CDS             /note="TTE1468"
                complement(76..1140)
                /gene="ELP3"
                /note="Best Blastp hit = gi|7462101|pir||A72295 conserved
                hypothetical protein - Thermotoga maritima (strain MSB8)
                gi|981648|gb|AAD36179.1|AE001769_9 (AE001769) conserved
                hypothetical protein [Thermotoga maritima], score 202,
                E-value 6.00E-51"
                /codon_start=1
                /transl_table=11
                /product="ELP3 component of the RNA polymerase II complex,
                consists of an N-terminal BioB/LipA-like domain and a
                C-terminal histone acetyltransferase domain (only
                BioB/LipA-like domain in Thermotoga)"
                /protein_id="AAM24690.1"
                /db_xref="GI:20516483"
                /translation="MKRKMIIPIFIPHWGCSFKCVFCNQNLITGQKKDVTVEEVIET
                IESHLKTIPEGSEVEVSFFGGSFTGIPIEKQNLYLSIAREYLKRGQIQGIRLSTRPDY
                ITEBILENLKKHDVSVIELGVQSMEDDVLEKSRGHTSEDVVNAVNLIRRYDPFKLGLQ
                LMIGLPGDTEKSLNTAHKIVSLKFDVRIYPTLVIRNTYLEKKYKEGKYTPLSLEQA
                INISKKMYIIFVKAGIDVIRIGLQTTDNINFNKDVVAGPFHPAMGQLVESSIMLDVLK
                EVIEKGIKGKKISIFCNERKLSTVIGQKKYNKRFLEDKYGLQVQVKLLEKIIEDKIV
                LSYDKKLYPISITDFIKDNP"

terminator       complement(1091..1108)
                /gene="ELP3"
                /note="putative rho-independent transcription terminator"

gene             complement(1137..1837)
                /gene="Rnc"
                /note="TTE1469"

CDS              complement(1137..1823)
                /gene="Rnc"
                /note="Best Blastp hit = gi|10175109|dbj|BAB06208.1|
                (AP001515) ribonuclease III [Bacillus halodurans], score
                193, E-value 1.00E-48"
                /codon_start=1
                /transl_table=11
                /product="dsRNA-specific ribonuclease"
                /protein_id="AAM24691.1"
                /db_xref="GI:20516484"
                /translation="MLTYLEQKINYEFKDKTLLEALITHSSWAHEGKNEKVSNERLEF
                LGDSVLSLVISEVLYKNRKDLEEGSLSKYRAEIVCEPSLARCARKIELGSFLRMGKGE
                EISGGRDRDSILADAMEALLAAVVLDGGLEAVRRVILDLFKEIIDEVLKGIIIRDYKT
                RLQEVVQSMEVGKTYELVEEIGPDHNKTFVTQVKIGDVVLGIGQGKSKKESQAAAM
                EALGKLGILK"

RBS              complement(1149..1153)
                /gene="ELP3"

misc_feature     complement(1152..1352)
                /gene="Rnc"
                /note="Pfam match to entry dsrm, Double-stranded RNA
                binding motif, score 85.2, E-value 1.30E-21"

misc_feature     complement(1434..1706)
                /gene="Rnc"
                /note="Pfam match to entry Ribonuclease_3, RNase3 domain.,
                score 139, E-value 8.40E-38"

RBS              complement(1833..1837)
                /gene="Rnc"

gene             complement(1849..3092)
                /gene="FabB"
                /note="TTE1470"

CDS              complement(1849..3078)
                /gene="FabB"
                /note="Best Blastp hit = gi|7433750|pir||G69842
                3-oxoacyl-[acyl-carrier-protein] synthase (EC 2.3.1.41) -
                Bacillus subtilis gi|2633471|emb|CAB12975.1| (Z99109)
                similar to 3-oxoacyl- acyl-carrier protein synthase
                [Bacillus subtilis] gi|2633468|emb|CAB12991.1| (Z99110)
                similar to 3-oxoacyl- acyl-carrier protein synthase
                [Bacillus subtilis], score 496, E-value 1.00E-139"
                /codon_start=1
                /transl_table=11
                /product="3-oxoacyl-(acyl-carrier-protein) synthase"
                /protein_id="AAM24692.1"
                /db_xref="GI:20516485"


                /translation="MNRVVVTGLGVISPLGNTVEKFWNSLINGENGIGLVTKFDTSEF
                PTKVAAEVKDFDPTLYIDKKEAARMDRFIQFAVASTKLALEDAGIELDKVDLNRFGVI
                YGTGIGGIETFENQMKVMYEKGPSKVSPFFIPMMIANMAAGQIAITFGKGINETIVN
                ACASSTNAIGEAFKAIQRGDADLIVTGGTEAAITPMSFAGFCAMKAMSTNPDPSTASR
                PFPLNRDGFVMGEGSATLILESLEHAQKRGAKIYAELIGYGATDDAYHITAPAPEGEG
                AARAMEAALKDGKVSYDMVDYINAHGTSTEYNDKFETMAIKKVFKDHAYKLCVSSNKS
                MGGHLLGAAGAVBAVATVLTITHGIVPPTINYQTPDPECDLNYVPNKAIEREVNYAIS
                NSFGFGGHNATLLFKKF"
misc_feature     complement(1855..2325)
                /gene="FabB"
                /note="Pfam match to entry ketoacyl-synt_C, Beta-ketoacyl
                synthase, C-terminal domain, score 238.8, E-value
                7.70E-68"

misc_feature     complement(2347..3078)
                /gene="FabB"
                /note="Pfam match to entry ketoacyl-synt, Beta-ketoacyl
                synthase, N-terminal domain, score 295.3, E-value
                7.70E-85"

RBS              complement(3088..3092)
                /gene="FabB"

terminator       complement(3111..3127)
                /note="putative rho-independent transcription terminator"

gene             complement(3133..3432)
                /gene="AcpP"
                /note="TTE1471"

CDS              complement(3133..3432)
                /gene="AcpP"
                /note="Best Blastp hit = gi|3912967|sp|O67611|ACP_AQUAE
                ACYL CARRIER PROTEIN (ACP) gi|7442262|pir||A70448 acyl
                carrier protein - Aquifex aeolicus
                gi|2984024|gb|AAC07567.1| (AE000752) acyl carrier protein
                [Aquifex aeolicus], score 75.5, E-value 2.00E-13"
                /codon_start=1
                /transl_table=11
                /product="Acyl carrier protein"
                /protein_id="AAM24693.1"
                /db_xref="GI:20516486"
                /translation="MLTEEWLCNIIISVQFLRGGEKGMIFEKVRNIIAEQLGIDPEEI
                TMESSFIDDLGADSLDIVELIMALEEEDIEIPDEDAEKIKTVGDVEYLSNIVE"

misc_feature     complement(3151..3354)
                /gene="AcpP"
                /note="Pfam match to entry pp-binding, Phosphopantetheine
                attachment site, score 95.2, E-value 9.80E-25"

gene             complement(3407..4163)
                /note="TTE1472"

CDS              complement(3407..4150)
                /gene="FabG3"
                /note="Best Blastp hit = gi|1502421|gb|AAC44307.1|
                (U59433) 3-ketoacyl-acyl carrier protein reductase
                [Bacillus subtilis], score 265, E-value 6.00E-70"
                /codon_start=1
                /transl_table=11
                /product="Dehydrogenases with different specificities
                (related to short-chain alcohol dehydrogenases)"
                /protein_id="AAM24694.1"
                /db_xref="GI:20516487"
                /translation="MVDEKRVAPVTGGSRGIGRAIAVRLAKDGFNIAITYVKDDKSAE
                EVVEEVKKGHGVDALAIKCDVSKYHEVEKAVEKVIEEFGSIDVVVNNAGITKDNLILKM
                EEEWDQVIDVNLKGAPNVIKFASKYMIKKRKGKINISSVVGLMGNVGQANYAASKA
                GIIGLIKSVAKBLASRGITVNAVAPGIETDMTNVLKEDIKEAMLKSIPLKRAGKPEE
                VAEVVAFLASSADYITGQVINVDGGMVM"

misc_feature     complement(3428..3520)
                /gene="FabG3"
                /note="Pfam match to entry adh_short_C2, Short chain
                dehydrogenase/reductase C-terminus, score 59.2, E-value
                3.40E-15"

misc_feature     complement(3578..4135)
                /gene="FabG3"
                /note="Pfam match to entry adh_short, short chain
                dehydrogenase, score 283.1, E-value 3.60E-81"

gene             complement(4154..4509)
                /gene="FabD"
```

```
CDS              /note="TEl473"
                 complement(4154..5083)
                 /gene="FabD"
                 /note="Best Blastp hit = gi|7404378|sp|P71019|FABD_BACSU
                 MALONYL COA-ACYL CARRIER PROTEIN TRANSACYLASE (MCT)
                 gi|7433733|pir||H69620 [acyl-carrier-protein]
                 S-malonyltransferase (EC 2.3.1.39) fabD [validated] -
                 Bacillus subtilis gi|2337819|emb|CAA74249.1| (Y13937)
                 putative FabD protein [Bacillus subtilis]
                 gi|2633962|emb|CAB13463.1| (Z99112) malonyl CoA-acyl
                 carrier protein transacylase [Bacillus subtilis], score
                 276, E-value 2.00E-73"

Alignment Scores:
Pred. No.:
Score:                 2.11e-50       Length:          11845
Percent Similarity:    775.50         Matches:         154
Best Local Similarity: 69.99%         Conservative:    58
Query Match:           50.83%         Mismatches:      90
DB:                    51.09%         Indels:          1
                       1              Gaps:            1

US-09-308-397-2 (1-306) x AE013105 (1-11845)

Qy     3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrTyrLeuGlyMetGlyArgAspPhe 22
Db  5080 AAAATAGCTTTTATATACCCGGCCAAGGTGCCCAGTACGCAGGGATGGGAAAGGAAATA 5021

Qy    23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db  5020 TACGAAAAATACGAGAAGCAGAAGAAATTTTGAAGAGCAGATGAGCTTTAGGTTTT 4961

Qy    43 AspLeuArgTyrLeu--IleAspThrGluGluLysLeuAsnGlnThrArgTyrThr 61
Db  4960 AACATATCTAAACTGTGCTTGAAGGGCCTGAAGAGGAGCTTATGAAGACGGAGAACACA 4901

Qy    62 GlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGlyTyr 81
Db  4900 CAGCCGGCCATTTTGACAGTTCAGTGGCCTTGACAAGGTTTTGCAGAAAAGAGGTGTA 4841

Qy    82 GlnProAspMetValAlaGlyTyrLeuSerLeuGlyGluTyrSerAlaLeuAlaSerGly 101
Db  4840 AAACCTGATGTGCACAGCAGGTTTAAGTTTAGGAGAGTATTCATCTCTCTGACTGGCAGAA 4781

Qy   102 AlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGlu 121
Db  4780 GCTTTAGATTTTGAGGATGCAGTGGAGACTTGTCAAAAAGAGAGAAAGTACATGCAGGAA 4721

Qy   122 AlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIle 141
Db  4720 GTGGTCCCAGAAGGAGTCCGGCACAATGGCAGCAATACTTGGACTTCCCAATAGGAAGTT 4661

Qy   142 GluGluLysGlnLysValAlaSerGluGluGlyValValThrProAlaAsnThrTyrAsnThr 161
Db  4660 GAGGAAAATATGCCGGATTCGATCCAGAGGTCGGTGTGGTAGAACCTGCTAACTACAACTGT 4601

Qy   162 ProLeuAlaGlnIleIleValIleAlaGlyGluValAlaAlaValAlaValAspArgArgValAlaValGluLeuLeu 181
Db  4600 CCGGGTCAATTAGTGGTCTCTGGGGAAGGTTAAGGCTGTAGAAGAGCCTGGAACTAGCG 4541

Qy   182 GlnGluAlaGlyAlaLysAlaArgLeuIleGluProLeuLysSerGlyProPheHisThrAla 201
Db  4540 AAGGAGAGGAGACTTAAGAAAGCCTAAGAAGAGCTAGTGCTTGCAGTAAGTGCCCTTTCCACTGCAGC 4481

Qy   202 LeuLeuGluProAlaSerGlnLysLysAlaGlyThrLeuValAlaGlnValSerPheSerAsp 221
Db  4480 ATGTTGAAAGAGCCTGGAGAACTTCTTAGCAAGGAATTGGATAAAGTAGAGATAAAGGAC 4421

Qy   222 PheThrCysProLeuValGlyAsnThrGlnAlaAlaValMetGlnLeuGluAspIleAla 241
Db  4420 TTGAAGGTTCCTGTGATATCAAATGTTACTGCCAATACGTGCAAAAAGATAAGGTTAAA 4361

Qy   242 GlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValValMet 261
Db  ...
```

(right column)
```
Db   4350 GAATTACTCATAAAACAGGTGAGTCATCCCGTCCTCTGGGAGCAGTCTGTGAGAAAATG 4301
Qy    262 GlnGluAlaGlyIleSerAsnPheIleGluIleGlyLeuGlyLysValLeuSerGlyPhe 281
Db   4300 ATAGAAGACGGTGTGGATACTTTCATAGAAATAAGGACCTGAAAGACTTTGAGCGGATTT 4241
Qy    282 ValLysLysIleAspGlnThrAlaHisLeuAlaHisGlueAspGlnAlaSerLeuVal 301
Db   4240 GTCAAGAGAAATTCATAGAAGCAGGACTGTTTGAATTTGAGGATGAGGAGTCTCCATG 4181
Qy    302 AlaLeuLeu 304
Db   4180 AAGGCTCTC 4172

RESULT 14
AE007854/c
LOCUS       AE007854              13165 bp    DNA     linear   BCT 27-JUL-2001
DEFINITION  Clostridium acetobutylicum ATCC824 section 342 of 356 of the
            complete genome.
ACCESSION   AE007854 AE001437
VERSION     AE007854.1  GI:15026666
KEYWORDS    .
SOURCE      Clostridium acetobutylicum.
  ORGANISM  Clostridium acetobutylicum
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 13165)
  AUTHORS   Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
            Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
            Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
            Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
  TITLE     Genome sequence and comparative analysis of the solvent-producing
            bacterium Clostridium acetobutylicum
  JOURNAL   J. Bacteriol. 183 (16), 4823-4838 (2001)
   MEDLINE   21359325
   PUBMED    11466286
REFERENCE   2  (bases 1 to 13165)
  AUTHORS   Childress,D., Zeng,Q. and Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2001) GTC Sequencing Center Production,
            Finishing, and Bioinformatics teams, Genome Therapeutics Corp., 100
            Beaver Street, Waltham, MA 02453-8443, USA
FEATURES             Location/Qualifiers
     source          1..13165
                     /organism="Clostridium acetobutylicum"
                     /strain="ATCC 824"
                     /db_xref="ATCC:824"
                     /db_xref="taxon:1488"
     gene            complement(74..499)
                     /gene="CAC3571"
                     /note="fabZ"
     CDS             complement(74..499)
                     /gene="CAC3571"
                     /codon_start=1
                     /transl_table=11
                     /product="Hydroxymyristoyl-(acyl carrier protein)
                     dehydratase"
                     /protein_id="AAK81494.1"
                     /db_xref="GI:15026667"
                     /translation="MSLSIEQIMBIIPHRYPMLLVDRVEIEPGKRAVGYKNVTFNEQ
                     IFQGHYGKRPIMPGVLMIEPALQGVAILSLDKYKGKKPILGAVVQNAKFRRMVVPGD
                     VLKLEIEIVKVKGPAGIGKGIATVNGEKAVBAEITFMIV"
     gene            complement(520..999)
                     /gene="CAC3572"
                     /note="accB"
     CDS             complement(520..999)
                     /gene="CAC3572"
                     /codon_start=1
                     /transl_table=11
                     /product="Biotin carboxyl carrier protein of acetyl-CoA
                     carboxylase"
                     /protein_id="AAK81495.1"
                     /db_xref="GI:15026668"
```

```
                    /translation="MDYKEIEELIKTIDNSGLTVLEIESNGTKITMRKGTSNEESTSV
                    VQNAQIQENKIERVVTTQTHDDKKVIDKPSVDDENIRVVTSPIVGTFYESSGPDSEA
                    FAKVGTQVKKGQTLCIIEAMKLMNEIBSEFDGEVVEVLVKNEEMVEYGQELFKIALK"
         gene       complement(1038..2273)
                    /gene="fabF"
         CDS        complement(1038..2273)
                    /gene="CAC3573"
                    /codon_start=1
                    /transl_table=11
                    /product="3-oxoacyl-(acyl-carrier-protein) synthase I"
                    /protein_id="AAK81496.1"
                    /db_xref="GI:15026669"
                    /translation="MNRRVVTGIGAVTPVGNNADSFWCSIKEGKCGIDKIKAPDATD
                    FKVKLAAEVKDFTPEDFIDKREANRMDRFSQFAIVAADEAIKDSKLDLESIDKNRFGV
                    IVGSGIGGIGTIEKQDEKLJTKGPGRVSPMTIPMIIANVASGNLAIRYGAKGICTTIV
                    TACASANNSIGBSFRNIKFGYSDVMISGGSEAGITPLSLAGFASMKAVTKSEDPKRAS
                    IPFDKDRSGFVMGEGSGIVILEBEHALKRGAKIYAEIVGYGATCDAYHITSPAPNGE
                    GGARAMKLAMEEDNVRPEDISYINAHGTSAYNDSFETCAIKTVLGEYAYKVPVSSTK
                    SMTGHLLGAGAVEAIICAKAIEEGFIPPTIGYKBADPEEDLDYVPNEGRNAEVNYVL
                    SNSLGPGGHNATLLFKKYK"
         gene       complement(2301..3050)
                    /gene="CAC3574"
                    /note="fabG"
         CDS        complement(2301..3050)
                    /gene="CAC3574"
                    /codon_start=1
                    /transl_table=11
                    /product="3-ketoacyl-acyl carrier protein reductase"
                    /protein_id="AAK81497.1"
                    /db_xref="GI:15026670"
                    /translation="MEKLLSGKVAVVTGAGRGLGRAIALKLAEEGANLVVNYRSSEAE
                    TQKLIKEIEELGSKAVAVKADISKYDEAETIIKKALDEYGTVDILVNNAGITKDNLLL
                    RMKEEDFDSVVNLKGAFNCIKHTSRVMLKKKSGKIINISSVIGLIGNAGQVNYAAA
                    KAGIIGMTKSVAKLLASRGITVNAVAPGIIKSDMTDALTDKQRESIVAAVPLNKVGEA
                    EDVANLVLFLASDLSSYITGQVINVDGGMVM"
         gene       complement(3060..3986)
                    /gene="CAC3575"
                    /note="fabD"
         CDS        complement(3060..3986)
                    /gene="CAC3575"
                    /codon_start=1
                    /transl_table=11
                    /product="Malonyl CoA-acyl carrier protein transacylase"
                    /protein_id="AAK81498.1"
                    /db_xref="GI:15026671"
                    /translation="MGKIAFVFSGGGSQYVGMGKDLYDNYQSAKETFDKADEVLGFKI
                    SELCFEGKDELNLITENTQPAVLITSIAALRALEEKGIKPDVVAGLLGEYSAHVCS
                    GSFSFEDAVRLVKKRGRYMQEAVPKGIGTMAAIGLEGDVVRGICAEQSKEGIVEVAN
                    YNCPQGIVIAGEVKAVEBSACTKLKESGARRTVMLSVSGFFHTSMLKSAAEKLEEELKN
                    INIEDMKVPVITNVTGDYVEDKDQIKGLLLKKQVMSSVRNEDTIRRMDDGVDTFIELG
                    PGKTLSSFIKKINRKMTIFNIEKAEDLNKIQL"
         gene       complement(3999..4931)
                    /gene="CAC3576"
         CDS        complement(3999..4931)
                    /gene="CAC3576"
                    /codon_start=1
                    /transl_table=11
                    /product="Dioxygenase related to 2-nitropropane
                    dioxygenase"
                    /protein_id="AAK81499.1"
                    /db_xref="GI:15026672"
                    /translation="MIKTQFCDIIGIKYPIIQGGWAWVADSSLAAGVSNAGGLGIIAA
                    ANAPVEYVRDEIRAKKLTDKPFGVNIMLLSDNAEEVAKMVCEEGVKVVTTGAGNPGK
                    YIDMKKEHDIKVIPVVASVALARRMERCGVDAVVAEGCSSGGHVGELITMALVPQVVD
                    TEVTGRSTGHPVRVLRNKLARKYKLMEKEGASPEEMEELGRGALPRAVREGDVDNGSV
                    MAGQIAGLINKEETCDEIVESMFKEAVEVIDRIK"
         gene       complement(4990..5220)
                    /gene="CAC3577"
                    /note="acp"
         CDS        complement(4990..5220)
                    /gene="CAC3577"
```

```
                    /codon_start=1
                    /transl_table=11
                    /product="Acyl Carrier Protein, ACP"
                    /protein_id="AAK81500.1"
                    /db_xref="GI:15026673"
                    /translation="MVLDKVRREVIANQISVDKETIDMATSFGKDLNVDSLDLFQIIIE
                    LEEKFDIQIEDTENIKTVGDAVKFIEAQLKKK"
         gene       complement(5263..6240)
                    /gene="CAC3578"
                    /note="fabH"
         CDS        complement(5263..6240)
                    /gene="CAC3578"
                    /codon_start=1
                    /transl_table=11
                    /product="3-oxoacyl-[acyl-carrier-protein] synthase III"
                    /protein_id="AAK81501.1"
                    /db_xref="GI:15026674"
                    /translation="MNSVEIIGTGSYVPEKIVTNEDMSKIVDTSDEWISSRTGIKERR
                    ISINENTDLGAKAALRAIENDNIKPERIDLIIVATTSPDSYTPSVACIVQEKIGAKN
                    AACPDLNAACTGFIFALNTASQFIKTGEYKTALVGTEVLSKILLDWQDRGTCVLFGDG
                    AGAVIIRGGDENGIIKACLGSDGTGKDFLHCPATNVINPFSDEKGLASSKISMNGREV
                    FKPAVKVMVSSVKVIEDSGLNIEBDIDYIVPHQANIRIIEFAAKKLGLSMDKFFINLQ
                    NYGNTSGATIPLAIDEMNKKGLLKRGAKIVVVGFGGGLTWGSMVLKWTK"
         gene       complement(6245..6709)
                    /gene="CAC3579"
         CDS        complement(6245..6709)
                    /gene="CAC3579"
                    /codon_start=1
                    /transl_table=11
                    /product="Transcriptional regulator, MarR/EmrR family"
                    /protein_id="AAK81502.1"
                    /db_xref="GI:15026675"
                    /translation="MDRITTINETLVELFNDILTIEQTALKQGGFNDLSVIEVHTIE
                    AIGMYVPRTMSEVAATLGTVGTLTTAVGNLVKKGYVKRERSEDRRVVKIALTKRGK
                    LAYRIHAKFHSDMVKEFVKGLSKEEERVLTSALEKLNRFFRDKYHIKNCKKD"
         gene       complement(6749..7816)
                    /gene="CAC3580"
         CDS        complement(6749..7816)
                    /gene="CAC3580"
                    /codon_start=1
                    /transl_table=11
                    /product="Dioxygenase related to 2-nitropropane
                    dioxygenase"
                    /protein_id="AAK81503.1"
                    /db_xref="GI:15026676"
                    /translation="MKLPLQIGDLKADIPIIQGGMGVGISRSSLASAVANCGVGTI
                    SGVQIGFDEPDFENTFNANIRALNKHIKKAKQESKRGIIAVNFMVAMNDYDKYVKSA
                    VDAGADLIVSGAGLPTALPKIVEGSVKKIAPIVSSPKAASVICKWMDKHGRIPDLIV
                    VEGPEAGGHLGFKAEQVKGSHEELSDIITGVLTAIKPEEKHGVKIPVVCGGGVFYGK
                    DIAKYIKLGASGVQMATRFIATEECDANLNYKNVYVNCKEDDIKIVKSPVGMPGRAIS
                    NKFVQSNDKNITITKCYNCLKPCNPKDTPYCISKALINAVNGNVDEALLFTGAKSYKI
                    DKIVKVKELINELISDAECEF"
         gene       complement(8050..8655)
                    /gene="CAC3581"
         CDS        complement(8050..8655)
                    /gene="CAC3581"
                    /codon_start=1
                    /transl_table=11
                    /product="HAD superfamily hydrolase"
                    /protein_id="AAK81504.1"
                    /db_xref="GI:15026677"
                    /translation="MIKNIVPDIGNVLLKFTPVEFLKSKYKDDKIVEALYENVFKSNE
                    WVELDRGTITEEEATCRFCKRDPHNAKYIREVMETWNEMFHVPVEGTSELMKELREKGY
                    KIYLLSNYHLKAFKLISDKYDFIRDVDGGIISSKVKLLKPELEIYKALTNSCGIKPEE
                    SIFIDDRRENVEAAIKLGFNGIVFKDAYELKGKFDELGIAI"
         gene       complement(8886..9440)
                    /gene="CAC3582"
         CDS        complement(8886..9440)
                    /gene="CAC3582"
                    /codon_start=1
```

Alignment Scores:
Pred. No.:   2.74e-45   Length: 13165

```
Score:                710.50        Matches:        146
Percent Similarity:   67.11%        Conservative:    56
Best Local Similarity: 48.50%       Mismatches:      96
Query Match:          1             Indels:           3
DB:                                 Gaps:             3
```

```
AP003189
LOCUS       AP003189              298050 bp    DNA     linear   BCT 10-JUL-2002
DEFINITION  Clostridium perfringens str. 13 DNA, complete genome, section 5/10.
ACCESSION   AP003189 BA000016
VERSION     AP003189.2  GI:18146727
KEYWORDS    .
SOURCE      Clostridium perfringens str. 13 (strain:13) DNA.
  ORGANISM  Clostridium perfringens str. 13
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE   1
  AUTHORS   Shimizu,T., Ohtani,K., Hirakawa,H., Ohshima,K., Yamashita,A.,
            Shiba,T., Ogasawara,N., Hattori,M., Kuhara,S. and Hayashi,H.
  TITLE     Complete genome sequence of Clostridium perfringens, an anaerobic
            flesh-eater
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (2), 996-1001 (2002)
   MEDLINE  21664373
    PUBMED  11792842
REFERENCE   2  (bases 1 to 298050)
  AUTHORS   Shimizu,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2001) Tohru Shimizu, Institute of Basic Medical
            Sciences, University of Tukuba, Department of Microbiology; 1-1-1
            Tennohdai, Tsukuba, Ibaraki 305-8575, Japan
            (E-mail:tshimizu@md.tsukuba.ac.jp, Tel:81-298-53-3354,
            Fax:81-298-53-3354)
COMMENT     On Jan 14, 2002 this sequence version replaced gi:18144662.
FEATURES             Location/Qualifiers
     source          1..298050
                     /organism="Clostridium perfringens str. 13"
                     /strain="13"
                     /db_xref="taxon:195102"
                     /note="anaerobic pathogen for gas gangrene"
     gene            complement(34..1056)
                     /gene="CPE1002"
     CDS             complement(34..1056)
                     /gene="CPE1002"
                     /note="340 aa, no significant homology."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB80708.1"
                     /db_xref="GI:18144663"
                     /translation="MSNKKINFKMNNSDKVKLINNSLVLDGISLKELSKKIDIQKKD
                     IIKFMNEEGFPFDSTDGYFKKDKEESTSIDIKPSVETSKSSLPKTPEREDVQSIIDSI
                     NEKLNIKDIETTNILLPEIKTSEKVSVTKEDNLEDNTKTSQSMTLEIPRVSRTARRKK
                     KKKTLIKRMISYFRNSFSKIFLIKKKENKSSSKIECNDNFNKKEVNIKSTNSTLEE
                     VAIEKLPVNDKIETTSHGLNSESKLPFEIKIPKNFKGRRRIPKVDCDLDEINFLNSND
                     YSLEDLKKDFRDLKLNSDFIKNINSDSLESNRDKSLEERVCNLEKKINELEEFFLEIF
                     QAMSNK"
     gene            1361..1540
                     /gene="CPE1003"
     CDS             1361..1540
                     /gene="CPE1003"
                     /note="59 aa, similar to gpu:AP001515_87 BH2353 gene
                     product from Bacillus halodurans (79 aa); 46.7% identity
                     in 45 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB80709.1"
                     /db_xref="GI:18144664"
                     /translation="MNIDELTKRINELHKKHKEEGLSEDEHKERELRKEYINRFKSN
                     LREQLKGIEPKNKKN"
     gene            1552..2475
                     /gene="hprK"
     CDS             1552..2475
                     /gene="hprK"
                     /note="307 aa, similar to gpu:AP001519_105 HPr (Ser/Thr)
                     protein kinase/phosphatase from Bacillus halodurans (310
                     aa); 45% identity in 282 aa overlap
                     CPE1004"
                     /codon_start=1
```

```
US-09-308-397-2 (1-306) x AE007854 (1-13165)

Qy      3 LysThrAlaPheLeuPheAlaGlyGlyAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe  22
            ||| :||||:||||:: ||::::||||||
Db   3980 AAGATTGCATTTGTTTTTCAGGACAGGGTTCGCACGTATGTTGGAATGGGAAAGGATCTA 3921

Qy     23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValGlyLeuGlyTyr  42
            ||::  ::| |||:::||||||||:::|
Db   3920 TACGATAATTACCAAAGCGCAAAGGAAAACTTTGATAAAGCTGATGAAGCTGGCTTT 3861

Qy     43 AspLeuArgTyrLeu---IleAspThrGluGluAspLysLeuAsnGlnThrHisArgTyrThr  61
            :: :  |||
Db   3860 AAAATCAGTGAATTATGCTTCGAAGGAAAAGATGAAGAATTAAATTAACTGAGAACACT 3801

Qy     62 GlnProAlaIleLeuAlaThrSerValAlaIleAlaIleTyrArgLeuLeu--GlnGluLysGly  80
            |||||:| |:
Db   3800 CAGCCAGCCAGTTTAATTACTAGTATAGCTGCTTAAGCACTTGAAGAAGAAAGCA 3741

Qy     81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAlaSer 100
            |||:    :||||||||||
Db   3740 ATAAAACCTGATGTTGTTGCAGGTCTAAGCTTAGGAGAGTATTCTGACATGTTTGCAGT 3681

Qy    101 GlyAlaLeuAspPheGluAspAlaValAlaLeuAlaLysArgGlyAlaTyrMetGlu 120
            |||::::||||||      :
Db   3680 GGGAGCTTTTCATTTGAGGATGCAGTAAGATAGTGAAAAAGAGGAGATATATGCA 3621

Qy    121 GluAlaAlaProAlaAlaAspSerGlyLysMetValAlaLeuAsnThrProValGluVal 140
            |||||   ||:::||
Db   3620 GAAGCAGTACCAAAAGGAATAGGTACTATGGCTACTATAATCGGACTTGAAGGTGATGTA 3561

Qy    141 IleGluGluAlaCysGlnLysLysSerGluGluLeuGlyValValThrProAlaAsnTyrAsn 160
            :::     |||
Db   3560 GTTAGAGGATATGGCGCTGAGCAATCAAAAGAAGAATAGTAGGTAGCAAATTATAAT 3501

Qy    161 ThrProAlaGlnIleLeuValIleAlaGlyGluValValAlaValAspAlaArgAlaValGluLeu 180
            ||| ::
Db   3500 TGTCCAGGACAAATTCTTATAGCCAGGAGAGTTAAAGCTGTTGAAAGCCTGTACTAAG 3441

Qy    181 LeuIleGluLysGlyAlaGlyLysAlaLysArgGluIleProLeuLysSerGlyValSerGlyProPheHisThr 200
            ||:::::||:       :::  :|
Db   3440 TTAAAGGAAAGCGGAGCAAGAAGAACAGTTTCAGTAAGTGGACCATTTCACACA 3381

Qy    201 AlaLeuGluGluProAlaSerGlnIleLysSerLeuAlaGluThrLeuAlaGlnIleValSerPheSer 220
            :::
Db   3380 TCTATGCTCTAAAGAGTGCTGCTGAAAAATTAGAAGAGAATTAAAATATAGAA 3321

Qy    221 AspPheThrCysProLeuValGlyValGlyAsn---ThrGluAlaAlaValMetGlnLysGlyGluAsp 239
            ||:::::|
Db   3320 GATAATGAAGTGACCAGTTATTACAAATGTAACAGGAGAGATAAGGATCAG 3261

Qy    240 IleAlaGlnLeuLeuThrArgGlnIleValLysGluValProValArgPheTyrGluSerIleGly 259
            ||||::||
Db   3260 ATAAAAGGTTTATTGAAAAAGCAAGTCATGAGCTCTGTTAGATGGGAAGATACAATAAGA 3201

Qy    260 ValMetGlnGluAlaGlyIleIleSerAsnPheIleGluGlyProGlyLysValLysValLeuSer 279
            ||:::
Db   3200 AGAATCATGGATGGATGGAGTAGATCACATTTATAGAGCTTGGACCAGGTAAGACCTTAAGT 3141

Qy    280 GlyPheValLysLysIleLeuAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSer 299
            |||:::::
Db   3140 TCATTTATAAGAAATAATAAAAATAGAAAAATGACTATATTTAATATTGAAAAAGCTGAAGAT 3081

Qy    300 Leu 300
            |||
Db   3080 TTA 3078
```

RESULT 15

```
          /transl_table=1
          /product="probable HPr Ser/Thr protein kinase/phosphatase"
          /protein_id="BAB80710.1"
          /db_xref="GI:18144665"
          /translation="MGVTIEKLLIKDFSLEVIQIGEENVPINVSDVNRPGLQLAGFPYNY
          FAPERLQVIGAEWSPLEDMSPDLRKKRLNKFFSPDISCLIITRGLEIHEELLKAARK
          RNLWILRSDMVTTKFISKTMYLSDKMAPETRLHGVLVDVYGIGMLITGESGIGKSET
          ALELIKRGHRLVTDDAVDIKEIDGDLIGRSPEITFGMLEVRGMGIIDVSALYGLSSIL
          NSKQIKIIHFEHWKDDGDYDRLGVNDEYQDILGVKVKKLRVPIRPGRNIAVIIEAAA
          ANYRYQRMSDISPVDIIEKRMLESMEKESKI"
gene      2520. .3449
          /gene="CPE1005"
CDS       2520. .3449
          /gene="CPE1005"
          /note="309 aa, similar to C-terminal of pir:S75981
          magnesium/cobalt transport protein sll0507 from
          Synechocystis sp. (strain PCC 6803) (387 aa); 22.4%
          identity in 263 aa overlap. 2 putative transmembrane
          regions were found by PSORT."
          /codon_start=1
          /transl_table=1
          /product="Probable divalent cation transporter"
          /protein_id="BAB80711.1"
          /db_xref="GI:18144666"
          /translation="MVIFNIEENFKVVEWKKDEFPAWILMEEDEVNLLEKMPSYECI
          NECIDYKQEQRIDCFKDYTFPLLINIEQVKQIISKELDIFLGKSFIVTVYKSNISLL
          YDLIEDIKEKNCQLFKVDKSPSIILYYILDRIIINNYNIISDLETKGDKIELEILKD
          PESKLNSLLFIRREVYKLRKLIKRYRYIGDTLVIDENEVIGEGNKVFKNINSKIEK
          LITSLDTLSQELAIVREAYSELANKTNELMKVFIITAVFLPLELITAIFSMSPESM
          PFKTCPYGFYILVTILSLMALILLLIFKIRKFL"
gene      3516. .4922
          /gene="CPE1006"
CDS       3516. .4922
          /gene="CPE1006"
          /note="468 aa, similar to pir:E70145 aminopeptidase I
          (yscI) homolog (EC 3.4.11.-) from Borrelia burgdorferi
          (458 aa); 45% identity in 456 aa overlap"
          /codon_start=1
          /transl_table=1
          /product="probable aminopeptidase I"
          /protein_id="BAB80712.1"
          /db_xref="GI:18144667"
          /translation="MKDVKVLKKEYENAWTKYEEADLKCVFSLSDRYREFMSVAKTER
          ECVKVLANMAESKGFKNFYEVIKNGEKVTAGDKLYSINMDKTITLIKVGSEPLENGLR
          IIGAHIDSPRMDVKQNPLYEDSGLALLDTHYYGGVKKYQWVTIPLAIHGVVVKKDGTR
          VDIKIGEDENDPVLGISDLLIHLSADQLDKKGAKVVEGEDLNILVGSMLKGTEKEA
          VKANILVLLNEKYGITEEDFVSAALEVVPAGKARDYGLRSMILAYGHDDRICAYTSA
          EALMDLENVDKTCVALLVDKBEIGSVGATGMQSRFFENIIAELMDRKGEYSELKIRRC
          LQNSMMLSADVTAAFDPNYPSVCBKKNTAYFGHGVVFSKYTGARGKAGCNANAEYIA
          HLRNIMDKNGVWQTGELGKVDQGGGGTIAYILACYNMEVIDCGVALQNMHAPLEVAS
          KADLIYBTKKCYKAFPEEA"
gene      5047. .5634
          /gene="CPE1007"
CDS       5047. .5634
          /gene="CPE1007"
          /note="195 aa, partially similar to gp:D90740_3
          Hypothetical 24.0K protein from Escherichia coli (476 aa);
          27.6% identity in 170 aa overlap. Putative N-terminal
          signal sequence and 1 putative transmembrane region were
          found by PSORT"
          /codon_start=1
          /transl_table=1
          /product="conserved hypothetical protein"
          /protein_id="BAB80713.1"
          /db_xref="GI:18144668"
          /translation="MKNVLKGNLIFLSFILLVTLFIERYNGENILLLSLIFYLEII
          FITRFPLKYINLGQAIKNRSIFSREHKNKNKFINKVLLEEKLDPLSNVFTRFBRKNNL
          EMWKGEYTLFPLIDIRKFSLINDALGRGKGKDEVLRBFGNGLNNISKSLASKNLIYGVYGG
          EEFLVAINSCNKDIIEVLNKLDNLKKIRLNSLNK"
gene      5707. .6570
          /gene="CPE1008"
CDS       5707. .6570
          /gene="CPE1008"
```

```
          /note="287 aa, similar to gp:BCE243712_8 YkoW protein from
          Bacillus cereus (892 aa); 30.9% identity in 278 aa
          overlap"
          /codon_start=1
          /transl_table=1
          /product="conserved hypothetical protein"
          /protein_id="BAB80714.1"
          /db_xref="GI:18144669"
          /translation="MFFAKERAKKEDFIYAIEYTEELRSIRKKGNNLIIRDIKDALINK
          KLYTVYQPQISSLDDKIKGYETLIRWNHETLGSISASDIVKTLEEIAIKDLDLYVFE
          STCEFQSRLIQNNIELQCSVNISINTLHGRGVSNLYKISKNVNLKPENITLELENA
          DLEVRGNAINEIKLKKYGFKISIDDFGKGYSSINRVLKIPPDQIKIPKDFISDIPNS
          KYIAIINSINFAKTLGABLVIEGVESKEYYDFFKLLDFDVIQGFYFSKALNEKEFLE
          YVNNLGVSVSI"
gene      6738. .7097
          /gene="mgsA"
CDS       6738. .7097
          /gene="mgsA"
          /note="119 aa, similar to sp:MGSA_BACSU METHYLGLYOXAL
          SYNTHASE (EC 4.2.99.11) (MGS) from Bacillus subtilis (137
          aa); 62.9% identity in 116 aa overlap
          CPE1009"
          /codon_start=1
          /transl_table=1
          /product="methylglyoxal synthase"
          /protein_id="BAB80715.1"
          /db_xref="GI:18144670"
          /translation="MKIALIAHDKKKEMIELAKDFEDKLSKHILVATGTTGLKIMQN
          TSLEVKRCKSGPLGGDQEIGAMVANHDVDMVIFLRDPLTAQPHEPDISALLRLCDVYK
          VPLATNTBSAKLIMADI"
gene      7278. .7613
          /gene="CPE1010"
CDS       7278. .7613
          /gene="CPE1010"
          /note="111 aa, no significant homology. Putative
          N-terminal signal sequence and 2 putative transmembrane
          regions were found by PSORT."
          /codon_start=1
          /transl_table=1
          /product="hypothetical protein"
          /protein_id="BAB80716.1"
          /db_xref="GI:18144671"
          /translation="MRGNINLQMSNWIILVSTLITIKYLENLIELCKPNKKIGLTTK
          KTIKKVIVFFILTVSIVIIGITFPYFKWRFLNAVSIGNRTLIVAFTQLFFGIIFTLF
          INRKKYKVK"
gene      7739. .7999
          /gene="CPE1011"
CDS       7739. .7999
          /gene="CPE1011"
          /note="86 aa, no significant homology. Putative N-terminal
          signal sequence and 1 putative transmembrane region were
          found by PSORT"
          /codon_start=1
          /transl_table=1
          /product="hypothetical protein"
```

```
Alignment Scores:
Pred. No.:                               298050
Score:                     5.12e-43      Length:          147
                           703.50        Matches:
Percent Similarity:        67.53%        Conservative:     61
Best Local Similarity:     47.73%        Mismatches:       97
Query Match:               46.34%        Indels:            3
Db:                        1             Gaps:              3

US-09-308-397-2 (1-306) x AP003189 (1-298050)

Qy     1 MetThrLysThrAlaPheLeuGluPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db 68429 ATGGCTAAACTAGGATTTTATTTGCAGGGCAAGGAGCCCAATATGTAGGAATGGGCAA 68488

Qy    21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
           :::
Db 68489 GAATTCTTGATAACTTGAAGAAGTAAAGAGGTATTTAAAAGAAGTAGTGAAGCCTTA 68548
```

```
Qy     41  GlyTyrAspLeuArgTyrLeuIleAspThrGlu---GluAspLysLeuAsnGlnThrArg  59
               ||||:::    |||:::           :|:::              ||||:::||:|||
Db  68549  GGCATAGATATGGGAAGAACTTTGTTTGTGATCCTGAGGGGATTATTAAATAAAACTGAA  68608

Qy     60  TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys  79
               ::||||||||||||||    |||:::||       |||                 :::
Db  68609  TTTACTCAGCCTGCAATAATAACCACAAACATGGCTATACTTACCAGCACTAGATAAATTA  68668

Qy     80  GlyTyrGlnProAspMetValAlaGlyLeuSerGlyTyrSerAlaLeuValAla  99
               |||                 ||||||||:::|||:::         |:::
Db  68669  GGAGTTAAATCTCATATAAGTTGTGATTAAGCTTAGGAGAATATTCTGACTTATTCAT  68728

Qy    100  SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet  119
               ||||:::|||              ||||||||||||||||||||||    :::|||
Db  68729  AGTGGTGCCATTAATTTGAAGAATGGAGTTAAACTAGTTAAAAAAGAGTAAATTCATG  68788

Qy    120  GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu  139
               :::|||||                 ||||||       |||||||||
Db  68789  CAAGAAGCTGTGGCAGAAGGTATTGTGGAATGGTTGCAGTTTTAAGAATGACTCCAGA  68848

Qy    140  ValIleGluGlnLysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyr  159
               :::||||||    ||||       |||:::|||
Db  68849  CAAGTTGATGAAATAATAGAAAAAGTTCCCATATGGAATAGTTTAAGAGTGCAAATTAT  68908

Qy    160  AsnThrProAlaGlnIleLeuValIleGlyValValAlaValAlaValAspArgAlaValGlu  179
               |||||||||||:::||||||||||||:::|||:::||
Db  68909  AATTCACCTGGCCAAATTGTTATTTCTGGAGAATTAGTAGCTTTAGAAAAAGCTATGGAA  68968

Qy    180  LeuLeuGlnGluAlaGlyLeuAlaLysArgLeuIleProLeuLysSerValSerGlyProPheHis  199
                                 ||||              |||||||||||||
Db  68969  TTTATAAAAGAAGTTGGTGGA---AGAGCTTATAAAATTACCTGTATCCATTTCAC  69025

Qy    200  ThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnIleValSerPhe  219
               :::|||||||||||:::|||:::                |||||:::|||
Db  69026  TGTTCAAGTTCACACCAGCAGCTGAGAATTAAATAAAATTCAATA  69085

Qy    220  SerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGlnLysGlu  238
               :::              |||        |||    :::       ||| :::|||
Db  69086  AATAAATTAAATGGAATTAGCTTAGTCTAATGTAAGGGAGGAGCCTATTTAGAAGATGAT  69145

Qy    239  AspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIle  258
               :::           ||||||:::                 |||
Db  69146  AATATAATAGAACTTTTAACTTCTCAAGTTAAAAAGCCTGTTTTATTTATAAATGATATA  69205

Qy    259  GlyValMetGlnLeuGlnAlaGlyIleSerAsnPheIleGluIleGlyLysProGlyLysValLeu  278
               :::|||:::                 ||||||||
Db  69206  GAAAAAATGATTGAAGTGGTGTAGATACTTTTATTGAAATAGGTCCTGGTAAAGCTTTA  69265

Qy    279  SerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAla  298
               |||||||||||::::::              |||:::
Db  69266  AGTGGATTTGTTAAAAAAATAATAAAAAATGTTACAGTGCTTAATGTTAGAGGATTTAAAA  69325

Qy    299  SerLeuValAlaLeuLeuGluGluLys  306
               ||||||||||
Db  69326  TCTTTAGAAAAAACATTATCAAAA  69349
```

Search completed: June 11, 2003, 21:02:50
Job time : 2443 secs

OM protein - nucleic search, using frame_plus_p2n model

Copyright (c) 1993 - 2003 Compugen Ltd.

```
Run on:        June 11, 2003, 20:05:52 ; Search time 245 Seconds
                                          (without alignments)
                                        2812.700 Million cell updates/sec

Title:         US-09-308-397-2
Perfect score: 1518
Sequence:      1 MTKTAFLFACGGAQYLGMGR........QTAHLAHVEDQASIVALLEK 306

Scoring table: BLOSUM62

               Xgapop 10.0 , Xgapext  0.5
               Ygapop 10.0 , Ygapext  0.5
               Fgapop  6.0 , Fgapext  7.0
               Delop   6.0 , Delext   7.0

Searched:      2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters:        4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09308397/runat_06062003_112348_28863/app_query.fasta_1.455
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09308397 @CGN_1_1_208 @runat_06062003_112348_28863 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     N_Geneseq_101002:*
     1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
     2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
     3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
     4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
     5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
     6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
     7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
     8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
     9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
    10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
    11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
    12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
    13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
    14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
    15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
    16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
    17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
    18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
    19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
    20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
    21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
    22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
    23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
    24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1518	100.0	921	19	AAV37239	DNA encoding a Fab
	2	1518	100.0	921	22	AAH90730	CFE 34 coding sequ
c	3	1518	100.0	19702	19	AAV52140	Streptococcus pneu
	4	1515	99.8	945	22	AAH90843	2CFE 34 coding seq
	5	1515	99.8	3580	19	AAV65243	S. pneumoniae gene
	6	1514	99.7	921	23	AAS55845	Streptococcus pneu
	7	1048	69.0	2365589	24	ABA90521	Genomic sequence o
	8	1036	68.2	942	24	ABN68641	Streptococcus poly
	9	996.5	65.6	924	24	ABN68640	Streptococcus poly
	10	996.5	65.6	2155561	24	ABN71527	Streptococcus poly
	11	937	61.7	942	23	AAS53186	Enterococcus faeca
	12	920	60.6	3656	20	AAX15514	Enterococcus faeca
	13	863	56.9	547	19	AAV37241	C-terminal coding
	14	858	56.5	1196	19	AAT98542	DNA encoding a S.
c	15	858	56.5	1199	18	AAX30801	Streptococcus pneu
c	16	858	56.5	1209	19	AAV37406	Streptococcus pneu
	17	858	44.2	636	18	AAX30880	Streptococcus pneu
	18	671.5	42.4	930	23	AAS52391	E. coli DNA for ce
	19	643.5	41.9	939	23	AAS53241	Haemophilus influe
c	20	636	41.9	1830121	17	AAT42063	Haemophilus influe
	21	627.5	41.3	950	24	ABK73852	Bacillus lichenifo
	22	627.5	41.3	960	24	ABK73795	Bacillus lichenifo
	23	613	40.4	450	19	AAV37240	N-terminal coding
	24	594	39.1	954	24	ABQ90247	N-terminal coding
	25	589.5	38.8	2301	24	ABQ70706	M. capsulatus gene
c	26	588.5	38.8	2944528	24	ABA03041	Listeria monocytog
	27	586	38.6	319630	24	ABQ67194	Listeria innocua c
c	28	586	38.6	3011208	24	ABQ69245	Listeria innocua D
	29	566	37.3	46593	21	AAA81456	N. meningitidis pa
	30	566	37.3	349980	21	AAF21612	Neisseria meningit
	31	566	37.3	837096	21	AAA81489	N. meningitidis pa
	32	556.5	36.7	906	23	AAS51823	Staphylococcus aur
	33	556.5	36.7	936	23	AAS54495	Staphylococcus aur
	34	556.5	36.7	939	23	AAS54159	Pseudomonas aerugi
	35	550.5	36.3	1170	21	AAZ51219	Staphylococcus aur
	36	547.5	36.1	927	19	AAV31440	Nucleotide sequenc
	37	547.5	36.1	999	22	AAD06204	Staphylococcus aur
	38	528	34.8	273254	21	AAC81914	Chlamydia pneumoni
c	39	528	34.8	1230025	20	AAX91990	Nucleotide sequenc
	40	515.5	34.0	936	24	ABN92524	Staphylococcus epi
	41	510	33.6	31940	22	AAF28526	Genomic fragment #
	42	479.5	31.6	41599	22	AAI66165	Bacillus subtilis
	43	474.5	31.3	6251	18	AAV74336	Staphylococcus aur
	44	454.5	29.9	1038602	20	AAZ01425	Complete genome se
	45	439.5	29.0	1357	21	AAC49929	Arabidopsis thalia

ALIGNMENTS

```
RESULT 1
AAV37239
ID   AAV37239 standard; DNA; 921 BP.
XX
AC   AAV37239;
XX
DT   24-SEP-1998  (first entry)
XX
DE   DNA encoding a FabD polypeptide.
XX
KW   FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine;
KW   screen; ss.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..921
FT                   /*tag= a
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

FT   /product= FabD
XX
PN   WO9822133-A1.
XX
PD   28-MAY-1998.
XX
PF   14-NOV-1997; 97WO-US20992.
XX
PR   18-NOV-1996; 96US-0031160.
XX
PA   (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI   Gentry DR, Lonsdale JT, Payne DJ, Pearson SC, Van Aller G;
XX
DR   WPI; 1998-312173/27.
DR   P-PSDB; AAW60856.
XX
PT   New isolated Streptococcus pneumoniae FabD gene - used to develop
PT   products for the diagnosis, prevention and treatment of bacterial
PT   diseases, particularly S. pneumoniae infection
XX
PS   Claim 6; Pages 5-6; 45pp; English.
XX
CC   The present sequence encodes a FabD protein of Streptococcus pneumoniae
CC   010093 (NCIMB 40801). The novel FabD polypeptides are related to other
CC   proteins of the malonyl-CoA:ACP family. The products can be used for the
CC   diagnosis of Streptococcus pneumoniae infections. Vectors containing
CC   the FabD DNA sequence can be administered directly to a mammal in
CC   order to produce the FabD peptide to provoke an antibody/T-cell response in
CC   order to prevent a disease. The peptide can be used to screen for
CC   compounds which modulate its activity.
XX
SQ   Sequence 921 BP; 247 A; 193 C; 230 G; 251 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 7.34e-148
                          1518.00   Length: 921
Percent Similarity: 100.00%   Matches: 306
Best Local Similarity: 100.00%   Conservative: 0
Query Match: 100.00%   Mismatches: 0
DB: 19   Indels: 0
                                               Gaps: 0

US-09-308-397-2 (1-306) x AAV37239 (1-921)

Qy   1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyAlaGlnTyrLeuGlyMetGlyArg 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGACTAAAACACGCCTTTTTATTTGCTGGTCAAGGTGCCCAAGTATCTAGGGATGGGACGG 60

Qy   21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspAspArgAlaSerGlnValLeu 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GATTTCTATGATCAGTATCCGATTGTCAAAGAAACCATTGATGACCGAGCGAGTCAGGTGCTC 120

Qy   41 GlyTyrAspLeuArgTyrLeuIleAspThrGluIleAspLysLeuAsnGlnThrArgTyr 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GGTTATGATTTACGTTATCTCATCGATACGGAAGAGATAGACAAACTCAATCAGACCCGCTAT 180

Qy   61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleAlaThrTyrArgLeuGlnGlyLysGly 80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCGTTACCGTTATTGCAAGAAAAAGGGC 240

Qy   81 TyrGlnProAspMetValAlaGlyGlyLeuSerLeuGlyTyrSerAlaLeuValAlaLeuValAlaSer 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 TATCAGCCTGATATGGTTGCTGGTGGTTTGTCTCTTGGAGAATACTCTGCCTTGGTGCAAGC 300

Qy 101 GlyAlaLeuAspPheGluAlaGlySerGlyLysMetGluAlaLysAlaArgGlyValAlaTyrMetGlu 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GGCGCGCTCTTGGATTTTGAAGATGGGGTTGCCTTGGAGCTAAGCGTGGAGCGGTATATGGAA 360

Qy 121 GluAlaAlaProAlaAlaAspSerGlyValLysMetGlyAlaValAlaValLeuAsnThrProValGluVal 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 GAAGCCGCTCCTGCTGACCTCTGGCAAGATGGTAGTGTTCAATACGCCAGTAGAGGTC 420

Qy 141 IleGluGluAlaCysGlnLysLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160

Db 421 ATTGAAGAGAGCCTGTCAAAAAGCTTCTGAACTTGGAGTGGTTACTCCAGCCAACTATAAC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 161 ThrProAlaGlnIleValIleGluAlaGlyValValAlaValAlaValAspValGluValGluLeu 180
Db 481 ACACCTGCACAAATCGTCATTGAGGCAGGTGTTGCAGTTGCGGTTGAACTT 540

Qy 181 LeuGlnGluAlaGlyAlaLysAlaArgLeuIleLeuProLeuLysSerValSerGlyProPheHisThr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 TTGCAAGAGGCAGGTGCCAAACGCTTGATCCTCTTAAGGTGTCAGGTCCCTTTCACACC 600

Qy 201 AlaLeuLeuGluProAlaSerGlnLysLeuLeuAlaGluThrLeuAlaGlnValSerPheSer 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 GCTCTCCCTTGAGCCTGCTAGCCAGAAACTTGCTGAAACCTCAGCTCAGTAAGTTTTCA 660

Qy 221 AspPheThrCysProLeuValGlyAsnThrGluLeuAlaAlaAlaValMetGlnLysGluAspIle 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGCTGATGCAAAAAGAGGACATT 720

Qy 241 AlaGlnLeuLeuThrArgGlnValLysGluValArgPheTyrGluSerIleGlyVal 260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 GCTCAGCTCCTTGACGCGTCAGGTCAAGGAACCCGTCGTTCGTTCATGAAAGTATTGGGGTC 780

Qy 261 MetGlnGluAlaGlyIleSerAsnPheIleLeuGlyProGlyLysValValLeuSerGly 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 ATGCAAGAGGCAGGCATAAGCAACTTTATCGAGATTGGACCGGGGAAAGTCTTGTCAGT 840

Qy 281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 TTTGTTAAAAAATTGATCAAGCTGCTCACTAGCTCATGTGAAGATCAAGCGAGTTA 900

Qy 301 ValAlaLeuLeuGluGluLys 306
        ||||||||||||||||||||
Db 901 GTAGCCACTTTTAGAAAAA 918

RESULT 2
AAH90730
ID   AAH90730 standard; DNA; 921 BP.
XX
AC   AAH90730;
XX
DT   02-OCT-2001 (first entry)
XX
DE   CFE 34 coding sequence.
XX
KW   Antibacterial; vaccine; gene therapy; bacterial cell wall viability;
KW   CFE; CEG; Conserved Essential Gene; bacterial infection;
KW   antisense therapy; antibiotic resistance; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200149721-A2.
XX
PD   12-JUL-2001.
XX
PF   29-DEC-2000; 2000WO-US35604.
XX
PR   30-DEC-1999; 99US-0174089.
XX
PA   (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI   Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Bruccoleri RE;
PI   Thanassi JA;
XX
DR   WPI; 2001-496721/54.
DR   P-PSDB; AAM01031.
XX
PT   Nucleic acids encoding conserved essential genes involved in bacterial
PT   replication which are potential targets for the treatment of antibiotic
PT   resistant bacterial infections -
XX
PS   Claim 8; Pages 202-203; 380pp; English.
XX

```
CC    The present invention relates to nucleic acids (AAH90701-AAH90918)
CC    encoding polypeptides (AAM01002-AAM01114), which are essential for the
CC    viability of a bacterial cell wall. The acronym CFE stands for "CFE For
CC    Expression", where CEG stands for "Conserved Essential Gene". The nucleic
CC    acids are useful for detecting the presence of proteins essential for the
CC    viability of a bacterial cell wall in samples such as cells, tissues,
CC    biological fluids, blood, serum, nose, ear or throat swabs with ligands,
CC    and for detecting corresponding target nucleic acid molecules with
CC    complementary sequences. The nucleic acids are also useful for
CC    determining whether a genomic nucleotide sequence of interest is
CC    essential for viability of a bacterial cell or whether it resides within
CC    an operon, by integrating an exogeneous nucleotide sequence comprising a
CC    portion of an open reading frame of the genomic sequence of interest
CC    (comprising 200-500 base pairs) into the genomic sequence of interest
CC    which confers a selectable phenotype to the cell, and determining cell
CC    viability with a selection agent such as chloramphenicol. The nucleic
CC    acids and proteins are also useful as vaccines and for treating bacterial
CC    infections with gene therapy and antisense therapy. The nucleic acids
CC    also enable identification of targets suitable for the treatment of
CC    antibiotic resistant bacterial infections.
XX
SQ    Sequence 921 BP; 246 A; 193 C; 232 G; 250 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                  7.34e-148    Length:       921
Score:                      1518.00      Matches:      306
Percent Similarity:         100.00%      Conservative:   0
Best Local Similarity:      100.00%      Mismatches:     0
Query Match:                100.00%      Indels:         0
DB:                         22           Gaps:           0
```

US-09-308-397-2 (1-306) x AAH90730 (1-921)

```
Qy      1 MetThrLysThrAlaPheLeuPheAlaAlaGlyAlaGlnTyrLeuGlyMetGlyArg     20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGACTAAAACAGCCTTTTTATTTGCTGGTGCAAGGTGCCCAGTATCTTAGGGATGGGGACGG     60

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu     40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATTTTACTTGTCCCCTAGTCGACAATAGAAGTGCTGAGCGAGTCAGGTGCTC    120

Qy     41 GlyTyrAspLeuArgTyrLeuIleLeuAspThrGluGluLysLeuLysAsnGlnThrArgTyr     60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGTTATGATTTGCGTTATCTCATCGATACGGAGAGGACAAAACTCAATCAGACCCGCTAT    180

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgGlnLeuGlnLysGlyGly     80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTATTGCAAGAAAAAGGGGC    240

Qy     81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer    100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TATCAGCCTGATATGGTCGCTGGTTTGTCTCTTGGAGAATACTCTGCCTTGGTGGCAAGC    300

Qy    101 GlyValAlaLeuAspPheGluAspPheAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu    120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGCGCCCTTGGATTTGAAGATGCGGTTGCCTTGGTAGCTAAGCGTGAGCCTATATGGAA    360

Qy    121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGluVal    140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAAGCGGCTCCTGCTGACTCTGGCAAGATGGTAGCAGTTGCATTGAATACGCCAGTAGAGGTC    420

Qy    141 IleGluGluAlaCysGlnLysGlnLysAlaSerGlnLeuGlyLeuGlyValValThrProAlaAsnTyrAsn    160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATTGAAGAAGCCTGCAAAAAGCTTCAAAAGCTTGGAGTGGTTACTCCAGCCAACTATAAC    480

Qy    161 ThrProAlaGlnIleValIleLeuAlaGlyValValAlaAlaValAlaAspArgAlaValGluLeu    180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGCAGTGGTTGAACTT    540

Qy    181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysGlnValSerGlyProPheHisThr    200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTGCAAGAGGCAGGTGCCAAACCGTTGATTCCCTTAAGGTGTCAGGTCCCCTTTCACACC    600
```

```
Qy    201 AlaLeuGluProAlaSerGlyGlyLysLeuAlaGluThrLeuAlaGlnValValSerPheSer    220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCTCTCCTTGAGCCAGCTAGCCAGAAACTAGCTGCAGAACTCAGCTCAGGTAAGTTTTCA    660

Qy    221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle    240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGCTGATCGAAAAGAGGACATT    720

Qy    241 AlaGlnLeuLeuThrArgGlnValLysGluProValAlaArgPheTyrGluSerIleGlyVal    260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCTCAGCTCTTGACGCGTCAGGTCAGGAACCCGTTCGTTTCTATGAAAGTATTGGGGTC    780

Qy    261 MetGlnGluAlaGlyIleSerArgAsnPheIleGluIleLeuGlyProGlyLysValLeuSerGly    280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ATGCAAGAAGCAGGCATAAGCAACTTATCGAGATTGGACCGGGGAAAGTCTTGTCAGGT    840

Qy    281 PheValLysIleAspGlnThrAlaHisLeuAlaHisValGlnAlaAspIleAlaSerLeu    300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TTTGTTAAAAAAATTGATCAACTGCTCACTTAGCTCATGTGGAGATCAAGCGAGTTTA    900

Qy    301 ValAlaLeuLeuGluLys    306
          ||||||||||||||||||
Db    901 GTAGCACTTTTAGAAAAA    918
```

```
RESULT 3
AAV52140/c
ID   AAV52140 standard; DNA; 19702 BP.
XX
AC   AAV52140;
XX
DT   23-OCT-1998 (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:7.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO981931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;    97WO-US19588.
XX
PR   31-OCT-1996;    96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 194-205; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
```

```
CC   isolating the nucleic acid molecules from the members; or (b) isolating
CC   mRNA, DNA or cDNA produced from an organism; amplifying nucleic acid
CC   molecules whose nucleotide sequence is homologous to amplification
CC   primers derived from the fragment of the S. pneumoniae genome to prime
CC   the amplification and isolating the amplified sequences. The computer
CC   readable medium can be used in a computer-based system for identifying
CC   fragments of the S. pneumoniae genome of commercial importance, or
CC   expression modulating fragments of the S. pneumoniae genome. Products
CC   from the present invention can be used in diagnosis kits and assays, and
CC   pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ   Sequence 19702 BP; 5736 A; 4548 C; 3728 G; 5690 T; 0 other;
```

```
Alignment Scores:
Pred. No.:
Score:                   1518.00          4.66e-146
Percent Similarity:      100.00%          Length:           19702
Best Local Similarity:   100.00%          Matches:          306
Query Match:             100.00%          Conservative:     0
DB:                                       Mismatches:       0
                                          Indels:           0
                                                            19
                                          Gaps:             0
```

```
US-09-308-397-2 (1-306) x AAV52140 (1-19702)

Qy        1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlyGlnTyrLeuGlyMetGlyArg 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17149 ATGACTAAAACAGCCTTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG 17090

Qy       21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValIleu 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17089 GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCCAGCCAGGTGCTC 17030

Qy       41 GlyTyrTyrAspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyr 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17029 GGTTATTATGATTTGCGTTATCTCATCGATACGGAAGAAGACAAACTCAATCAGACCCGCTAT 16970

Qy       61 ThrGlnProAlaIleAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16969 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC 16910

Qy       81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuLeuValAlaSer 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16909 TATCAGCCTGATATGGTTGCTGGTTTGTCTCTTGGATACTCTGCCTTGGTGGCAAGT 16850

Qy      101 GlyAlaLeuAspPheGluAspSerGlyLysGlyAlaTyrAlaTyrMetGlu 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16849 GGCGCCCTTGGATTTTGAAGATAGCGGTTGCCTTGGTAGCTAAGCCGTTGGAGCCTATATGGAA 16790

Qy      121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16789 GAAGCGGCTCCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGGCCAGTAGAGGTC 16730

Qy      141 IleGluGluAlaCysGlnLysAlaSerGlyLeuGlyValValThrProAlaAsnTyrAsn 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16729 ATTGAAGAAGCTTGCCAAAAGGCTTCAGGTCTTGGTGTTGTTACTCCAGCCAACTATAAC 16670

Qy      161 ThrProAlaGlnIleLeuValIleAlaGlyIleGluValAlaValAspArgAlaValGluLeu 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16669 ACACCTGCACAAATCCTCATTGCTGGAATTGGTGTTGCAGTTGATCGAGCGGTTGAACTT 16610

Qy      181 LeuGlnGluAlaGlyAlaLysAlaArgLeuIleProLeuValSerGlyValProPheHisThr 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16609 TTGCAAGAAGCAGGTGCCAAACGCTTAATTCCTCTTGTGTCAGGTGTCCCATTTCACACC 16550

Qy      201 AlaLeuLeuGlyLeuProAlaSerGlnLeuProLysAlaGluThrLeuValAlaGlnValSerProPheSer 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16549 GCTCTCCTTGGACCTGCTAGCCAGAAACTAGCTGCGAAACTCAGCTAAGTTTTCA 16490

Qy      221 AspPheThrCysProLeuValGlyLysThrGluAlaAlaValMetGlnLysGluAspIle 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16489 GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGCTGTGATGCAAAAAGAGGACATT 16430

Qy      241 AlaGlnLeuThrArgGlnValLysGluProPheTyrArgPheGluSerIleGluValVal 260
```

```
Db    16429 GCTCAGCTCTTGACGCGTCAGGTCAGGTCAAGGAACCCGTTCGTTCTCTATGAAAGATTGGGTC 16370
Qy      261 MetGlnGluAlaGlyIleSerAsnPheIleGluGlyProGlyLysValLeuSerGly 280
Db    16369 ATGCAAGAAGCAGGCATAAGCAACTTATCGAGATTGACCGGGAAGTCTTGTCAGGT 16310
Qy      281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
Db    16309 TTTGTTAAAAAAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCCAGTTTA 16250
Qy      301 ValAlaLeuLeuGluLys 306
Db    16249 GTAGCACTTTTAGAAAAA 16232
```

```
RESULT 4
AAH90843
ID   AAH90843 standard; DNA; 945 BP.
XX
AC   AAH90843;
XX
DT   02-OCT-2001 (first entry)
XX
DE   2CFE 34 coding sequence.
XX
KW   Antibacterial; vaccine; gene therapy; bacterial cell wall viability;
KW   CFE; CEG; Conserved Essential Gene; bacterial infection;
KW   antisense therapy; antibiotic resistance; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200149721-A2.
XX
PD   12-JUL-2001.
XX
PF   29-DEC-2000; 2000WO-US35604.
XX
PR   30-DEC-1999; 99US-0174089.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Dougherty TJ,  Pucci MJ,  Dougherty BA,  Davison DB,  Bruccoleri RE;
PI   Thanassi JA;
XX
DR   WPI; 2001-496721/54.
XX
PT   Nucleic acids encoding conserved essential genes involved in bacterial
PT   replication which are potential targets for the treatment of antibiotic
PT   resistant bacterial infections -
XX
PS   Claim 8; Fig 55; 380pp; English.
XX
CC   The present invention relates to nucleic acids (AAH90701-AAH90918)
CC   encoding polypeptides (AAM01002-AAM01114), which are essential for the
CC   viability of a bacterial cell wall. The acronym CFE stands for "CEG For
CC   Expression", where CEG stands for "Conserved Essential Gene". The nucleic
CC   acids are useful for detecting the presence of proteins essential for the
CC   viability of a bacterial cell in samples such as cells, tissues,
CC   biological fluids, blood, serum, nose, ear or throat swabs with ligands,
CC   and for detecting corresponding target nucleic acid molecules with
CC   complementary sequences. The nucleic acids are also useful for
CC   determining whether a genomic nucleotide sequence of interest is
CC   essential for viability of a bacterial cell or whether it resides within
CC   an operon, by integrating an exogeneous nucleotide sequence comprising a
CC   portion of an open reading frame of the genomic sequence of interest
CC   (comprising 200-500 base pairs) into the genomic sequence of interest
CC   which confers a selectable phenotype to the cell, and determining cell
CC   viability with a selection agent such as chloramphenicol. The nucleic
CC   acids and proteins are also useful as vaccines and for treating bacterial
CC   infections with gene therapy and antisense therapy. The nucleic acids
CC   also enable identification of targets suitable for the treatment of
CC   antibiotic resistant bacterial infections.
XX
SQ   Sequence 945 BP; 254 A; 205 C; 232 G; 254 T; 0 other;
```

Alignment Scores:
Pred. No.:
Score: 1.56e-147
Percent Similarity: 100.00%
Best Local Similarity: 99.67%
Query Match: 22
DB:

```
Length:         945
Matches:        305
Conservative:   1
Mismatches:     0
Indels:         0
Gaps:           0
```

US-09-308-397-2 (1-306) x AAH90843 (1-945)

```
Qy    1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGACTAAAACAGCCTTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG  60

Qy   21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGTCAGGTGCTA 120

Qy   41 GlyTyrTyrAspLeuArgTyrLeuIleAspThrGluLysLysLeuAsnGlnThrArgTyr  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGTTATTATGATTTGCGTTATCTCATCGATACGGAAGAAGACAAACTCAATCAGACCCGCTAT 180

Qy   61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly  80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 ACGCAACCAGCCATTCTAGCACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC 240

Qy   81 TyrGlnProAspMetValAlaGlyTyrLeuLeuSerLeuGlyTyrSerAlaLeuValAlaSer 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TATCAGCCTGATATGGTTGCTGGTTGTCTCTTTGGAAATACTCTGCCTTGGTGGGCAAGC 300

Qy  101 GlyAlaLeuAspPheGluAspProAlaValAlaLeuAlaLeuAlaLysArgGlyAlaTyrMetGlu 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GGGCGCCTTGGATTTGAAGATGCGGTTGCCCTTGGCTAGCTAAGCGTGGAGCCTATATGGAA 360

Qy  121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GAAGCGGCTCCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC 420

Qy  141 IleGluGluValCysGlnLysAlaSerGluGluGlyValValThrProAlaAsnTyrAsn 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATTGAAGAAGAGCCTGTCAAAAAGCTTCGAACTTGGAGCGTGGTTACTCCAGCCAACTATAAC 480

Qy  161 ThrProAlaGlnIleValIleAlaGlyValGluValAlaValAlaValAspArgAlaValGluLeu 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGCAGTTGATCGAGCGGTTGAACTT 540

Qy  181 LeuGlnGluAlaGlyAlaLysArgLeuIleLeuProLeuLysValSerGlyProPheHisThr 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TTGCAAGAAGCAGGTGCCAAACGCTTGATTCTCCTATAAGGTGTCAGGTCCCTTTCACACC 600

Qy  201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
        :|||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TCCTCCCTGAACCTGCTAGCCAGAAACTAGCTGAAACTCTGGCTCAGGTAGTTTTTTCA 660

Qy  221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaAlaValMetGlnLysGluAspIle 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGCTGTGATGCAAAAAGAGGACATT 720

Qy  241 AlaGlnLeuThrArgGlnValLysGlyProValArgPheTyrGluSerIleGlyVal 260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GCTCAGCTTACCCGTCAGGTCAAGGAACCCGTTCGTTTCTATGAAAGTATTGGGGTC 780

Qy  261 MetGlnGluAlaGlyIleIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 ATGCAAGAAGCAGGCATAAGCAACTTTATCGAGATTGGACCGGGGAAAGTTTGTCAGGT 840

Qy  281 PheValLysIleIleAspThrGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 TTTGTAAAAAATTGATCAAACTCAGACTGCACATTTAGCTCATGTGGAAGATCAAGCGAGTTTA 900

Qy  301 ValAlaLeuLeuGluLys 306
        ||||||||||||||||||
```

```
Db    901 GTAGCACTTTAGAAAA 918
```

RESULT 5
AAV65243
ID   AAV65243 standard; DNA; 3580 BP.
XX
AC   AAV65243;
XX
DT   24-DEC-1998 (first entry)
XX
DE   S. pneumoniae gene fragment.
XX
KW   Streptococcus pneumoniae protein; recombinant; gene expression;
KW   DNA chip; virulence; antibody; infection; detection; treatment; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9826072-A1.
XX
PD   18-JUN-1998.
XX
PF   09-DEC-1997; 97WO-US22578.
XX
PR   13-DEC-1996; 96US-0036281.
XX
PA   (ELIL) LILLY & CO ELI.
XX
PI   Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
PI   Mills BJ, Norris FH, Peery RB, Rockey PK, Rosteck PR;
PI   Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;
PI   Young Bellido ML;
XX
DR   WPI; 1998-348529/30.
XX
PT   Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
PT   for evaluating gene expression, and identification of virulence
PT   genes
XX
PS   Claim 1; Pages 96-98; 333pp; English.
XX
CC   This sequence is a gene fragment from the Streptococcus pneumoniae
CC   genome. The invention provides DNA sequences (AAV65201 to AAV65304) from
CC   the Streptococcus pneumoniae genome and corresponding protein sequences
CC   (AAW80605 to AAW80728). A recombinant host containing a vector comprising
CC   any of the above nucleic acids can be used for the recombinant expression
CC   of the protein sequences. The invention also provides a DNA chip having
CC   arrayed on it at least 15 base pair fragment of any one or more of these
CC   DNA sequences. The DNA chip can be used methods for evaluating gene
CC   expression in S. pneumoniae and for identifying virulence genes in
CC   S. pneumoniae. Antibodies that selectively bind to the above proteins or
CC   peptide fragments can be used to treat S. pneumoniae infection. The
CC   antibodies can also be used to detect S. pneumoniae cells.
XX
SQ   Sequence 3580 BP; 1035 A; 666 C; 894 G; 985 T; 0 other;

Alignment Scores:
Pred. No.:
Score: 9.45e-147
Percent Similarity: 1515.00
Best Local Similarity: 100.00%
Query Match: 99.67%
DB: 99.80%

```
Length:          3580
Matches:         305
Conservative:    1
Mismatches:      0
Indels:          0
Gaps:            19
```

US-09-308-397-2 (1-306) x AAV65243 (1-3580)

```
Qy    1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  718 ATGACTAAAACAGCCTTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG 777

Qy   21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  778 GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGTCAGGTGCTA 837
```

```
Qy   41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyr 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  838 GGTTATGATTTGCGTTACTCATCGATACGGAGAGACAAACTCAATCAGACCCGCTAT 897

Qy   61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleThrArgLeuLeuGlnGlyLysGly 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  898 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCACCGTTATTGCAAGAAAAGGGC 957

Qy   81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyValTyrSerAlaLeuValAlaSer 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  958 TATCAGCCTGATATGGTTGCTGGTTTGTCTCTTGGAGAATACTCTGCCTTGGTGGCAAGC 1017

Qy  101 GlyAlaLeuAspPheGluAspAlaValAlaAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1018 GGCGCCTTGGATTTTGAAGATGCGGTTGCCTTGGTAGCGTGGAGCCTATATGGAA 1077

Qy  121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1078 GAAGCGGCGCCTCCTGCTGACTCCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC 1137

Qy  141 IleGluGluAlaCysGlnIleAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1138 ATTGAAGAAGCCTGTCAAATAGCTTCGAACTTGGAGTGGTTACTCCAGCCAACTATAAC 1197

Qy  161 ThrProAlaGlnIleValIleLeuAlaGlyGluValValAlaValAlaAspArgAlaValGluLeu 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1198 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGATCGAGCGGTTGAACTT 1257

Qy  181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysSerGlyValSerGlyProPheHisThr 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1258 TTGCAAGAAGCAGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCAGGTGTTCCCTTTCACACC 1317

Qy  201 AlaLeuLeuGluProAlaSerGlnLysValThrLeuAlaGlnThrLeuAlaGlnIleSerPheSer 220
         ::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1318 TCTCTCCTTGAACCTGCTAGCCAGAAAGTTACCCTGGCTCAGACTCTGGCTCAGATTTTTTCA 1377

Qy  221 AspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIle 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1378 GATTTTACTTGTCCCCTAGTGGGCAATACAGAAGCTGCTGTGATGCAAAAAGAGGACATT 1437

Qy  241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1438 GCTCAGTTGTTCTTCACGCGGCGTCAGTCAGGTCAAGGAACCGGTTCGTTTCTATGAAAGTATTGGGGTC 1497

Qy  261 MetGlnGluAlaGlyIleSerAspIleSerAsnPheIleGlyIleGlyProGlyLysValLeuSerGly 280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1498 ATGCAAGAAGCAGGCATAAGCAACCACTTTATCGAAGATTTGAACCGGGAAAGTCTTGTCAGGT 1557

Qy  281 PheValLysLysAspGlnIleLysAspGlnThrAlaHisIleLeuAlaHisSerValGluAlaGlnSerLeu 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1558 TTTGTAAAAAAATTGATCAAACTGCTCACTTAGCTCACATGTGGAAGATCAAGCCAGTTTA 1617

Qy  301 ValAlaLeuGluLys 306
         |||||||||||||||
Db 1618 GTAGCACTTTTAGAAAA 1635
```

RESULT 6
AAS55845

```
ID   AAS55845 standard; DNA; 921 BP.
XX
AC   AAS55845;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Streptococcus pneumoniae DNA for cellular proliferation protein #416.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200170955-A2.
XX
```

```
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI   Yamamoto RT, Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU37986.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 9482; 51pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 921 BP; 246 A; 191 C; 232 G; 252 T; 0 other;
```

Alignment Scores:

Pred. No.:		Length: 921
Score: 1.91e-147		Matches: 305
Percent Similarity: 100.00%		Conservative: 1
Best Local Similarity: 99.67%		Mismatches: 0
Query Match: 99.74%		Indels: 0
DB: 23		Gaps: 0

US-09-308-397-2 (1-306) x AAS55845 (1-921)

```
Qy   1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGACTAAAACAGCCTTTTTATTTGCTGGTCAGGTGCCCAGTATCTAGGGATGGGACGG 60

Qy  21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 GATTTCTATGATCAGTATCCGATTGTCAAAGAAACGATTGATCGAGCGAGTCAGGTGCTA 120

Qy  41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyr 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GGTTATGATTTGCGTTATCTCATCGATACGGAGGAGACAAACTCAATCAGACCCGCTAT 180

Qy  61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleThrArgLeuLeuGlnGlyLysGly 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCACCCGTTATTGCAAGAAAAAGGGC 240
```

```
Qy    81  TyrGlnProAspMetValAlaGlyLysSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  TATCAGCCTGATATGGTTGCTGGTTTGTCTCTTGGAGAATACTCTGCCTTGGTGGCAAGC 300

Qy   101  GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GGGCGCCTGGATTTTGAAGATGCGGTTGCCCTGGTAGCTAAGCGTGGAGCCTATATGGAA 360

Qy   121  GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GAAGCGGCTCCTGCTGACTCTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTC 420

Qy   141  IleGluGluAlaCysGlnLysAlaSerGluLeuGlyLeuGlyValValThrProAlaAsnTyrAsn 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ATTGAAGAAGCCTGTCAAAAAGCTTCTGAACTTGGAGGTGGTTACTCCAGCCAACTATAAC 480

Qy   161  ThrProAlaGlnIleValIleAlaGlyValGluValAlaAlaValAspArgArgValAlaGluLeu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTTGATCGAGCCGTTGAACTT 540

Qy   181  LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  TTGCAAGAAGCAGGTGCCAAACGTTGATCCTTAAGGTGTCAGGTCCCTTTCACACC 600

Qy   201  AlaLeuLeuGluProAlaSerArgGlnLysGlyAlaThrLeuAlaGlnValSerPheSer 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  GCTCTCCTTGAACCTGCTAGCCAGGAACCAGTTGAAACTCAGCTCAGGTAAGTTTTTCA 660

Qy   221  AspPheThrCysProLeuValGlyAsnThrGluAlaAlaAlaValMetGlnLysGluAspIle 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGCTGTGATGCAAAAAGAGGACATT 720

Qy   241  AlaGlnLeuLeuThrArgGlnValValLysGluProValArgPheTyrGluSerIleGlyVal 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  GCTCAGCTCCTTGACGCGTCCAGGTCAAGGAACCCGTTCGTTTCTTATGAAAGTATTGGGGTC 780

Qy   261  MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  ATGCAAGAAGCAGGCATAAGCAACTTATCGAGATTGGACCCGGGAAAGTTTGTCAGGT 840

Qy   281  PheValLysLysIleAspGlnThrArgAlaHisLeuAlaHisValGlyAspAspProAlaSerLeu 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  TTTGTTAAAAAAATTGATCAGACTCGCTCACTTAGCTCTCATGTGGAAGATCAAGCGAGTTTA 900

Qy   301  ValAlaLeuGluLys 306
              |||||||||||||||
Db   901  GTAGCACTTTTAGAAAAA 918
```

RESULT 7
ABA90521
ID   ABA90521   standard; DNA; 2365589 BP.
XX
AC   ABA90521;
XX
DT   16-MAY-2002   (first entry)
XX
DE   Genomic sequence of Lactococcus lactis IL1403.
XX
KW   Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
XX
OS   Lactococcus lactis IL1403.
XX
FR   FR280746-A1.
XX
PD   12-OCT-2001.
XX
PF   11-APR-2000; 2000FR-0004630.
PR   11-APR-2000; 2000FR-0004630.
XX
PA   (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Bolotine A,   Sorokine A,   Renault P,   Ehrlich SD;

XX   WPI; 2002-04318/06.
DR
XX
PT   New nucleotide sequence useful in the identification or Lactococcus
PT   lactis and related species -
XX
PS   Claim 1; SEQ ID 1; 2504pp; French.
XX
CC   The present invention is related to a Lactococcus lactis nucleotide
CC   sequence (ABA90521) and related proteins (ABS3300-ABB55621).The
CC   nucleic acid sequence is useful in the detection and/or amplification of
CC   nucleic acid sequence, particularly to identify Lactococcus lactis or
CC   related species.The proteins of the invention are useful for the
CC   biosynthesis or biodegradation of a composition of interest. The
CC   invention helps research in lactic bacteria, particularly useful in the
CC   production of yogurt and cheese.
CC   Note: The sequence data for this patent is based on equivalent patent
CC   WO200177334 (published 18-OCT-2001) which is available in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

Alignment Scores:
Pred. No.:   Length: 2365589
Score: 1048.00   Matches: 208
Percent Similarity: 80.52%   Conservative: 40
Best Local Similarity: 67.53%   Mismatches: 58
Query Match: 69.04%   Indels: 2
DB: 24   Gaps: 1

US-09-308-397-2 (1-306) x ABA90521 (1-2365589)

```
Qy     1  MetThrLysThrAlaPheLeuPheAlaGlyValGlnGlyAlaGlyAlaGlnTyrLeuGlyMetGlyArg 20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 782872  ATGACTAAAACACAGCATTTTATTCTCCAGGTCAAGGGCACAAAAGCTTGGAATGGCCACGT 782931

Qy    21  AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValValLeu 40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 782932  GACTTATATGACCAATATGAAACAGTTAAAGCAACTTTGATGAAGCAAGTCAAGCTTTA 782991

Qy    41  GlyTyrAspLeuArgTyrLeuIleAspThrGluLeuAspLysLeuAsnGlnThrArgTyr 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 782992  GGATATGATTTGCGAGCTTTGATTGATAATGATGAAGAAACTAAATGAAACAAAGTAC 783051

Qy    61  ThrGlnProAlaIleLeuAlaThrSerValAlaIleThrTyrArgLeuLeuGlnGluLysGly 80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 783052  ACTCAACCTGCAATTTTAGCAACCTCTGTTGCTATTTTACGTTTGTTAAGTGAGATGGG 783111

Qy    81  TyrGlnProAspMetValAlaGlyLysLeuGlyGluTyrSerAlaLeuValAlaSer 100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 783112  ATTAAACCTGACCTTGTTGCTGGTCTTGGTGAATATTCTGCCTTGGTAGCATCA 783171

Qy   101  GlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGlu 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 783172  GGAATCATTGATTTTCAAGAGACAGTTAAGCTGTAGCTAAACGTGGTCAATATATGACA 783231

Qy   121  GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
             |||||||||||||| ||||||||||||||||||| :::
Db 783232  GAAGCTGCACCGACTGGTTCTCGGTAAAATGGTCGCTGTAATGAAACACAGACCCAGCCTTG 783291

Qy   141  IleGluGluAlaCysGlnLysAlaSerGluLeu------GlyValValThrProAlaAsn 158
             |||||||||||||||||||||||||||::::|||| :::||||||||:::
Db 783292  ATTGAAGAAAATTGCCAGAAAGCGGCCGAATTAAAGCGTATTGGAGTCCAGCAAAT 783351

Qy   159  TyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgArgValAlaVal 178
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 783352  TATAACACGTCCCGACACAAATGTTATTGGTGGTGAAGTTGAGGCCGGTTGAGTATGCTGTT 783411

Qy   179  GluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPhe 198
             |||||||||:::::||||||||||||||||||||||||||||::::
Db 783412  GAGTTGCTAAAAGAAGCCGAGTTCGTAAACTTATTGAATTAAAAGTTTCAGGACCTTTC 783471

Qy   199  HisThrAlaLeuLeuGluProAlaSerGlnLysGlyAlaThrLeuAlaGlnValSerValSer 218
```

CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic reagent. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 942 BP; 285 A; 159 C; 217 G; 281 T; 0 other;

Alignment Scores:
Pred. No.:   Length: 942
Score: 6.63e-98   Matches: 1036.00
Percent Similarity: 81.05%   Conservative: 39
Best Local Similarity: 68.30%   Mismatches: 58
Query Match: 68.25%   Indels: 0
DB: 24   Gaps: 0

US-09-308-397-2 (1-306) x ABN68641 (1-942)

Qy   1 MetThrLysThrAlaPheLeuPheAlaGlyAlaGlnTyrLeuGlyAlaGlnTyrLeuGlyMetGlyArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7 ATGACAAAGACAGCCTTTTTATTTGCCGGTCAAGGTGCTCAAGGTGCTCAAAAATTAGGGATGGCAGG 66

Qy   21 AspPheTyrArgGlnTyrProLeuValLysGluThrIleAspArgAlaSerGlnValLeu 40
        ||| :::   ::::| ||||||||:||| :::::
Db   67 GATTTTTATGATAAACTTGCTATTGTAAGAAAAACCTTTGATCAAGCTAGTCAAGTATTG 126

Qy   41 GlyTyrAspLeuArgTyrLeuIleAspThrGluLysAspLysLeuAsnGlnThrArgTyr 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 127 GGATACGATTTGCGCGTTTGATTGATACGTGACGAGTAAAACTTAATCAAACTAGCTAT 186

Qy   61 ThrGlnProAlaIleLeuLeuAlaThrSerValAlaIleAlaTyrArgLeuLeuGlnGluLysGly 80
        |||||||||||||||||||||||||||| :::||||||||:::
Db 187 ACGGAACCAGCTATTTTGCATCATCAATTGCATTACCGTGTTTGGGCTTACATCAC 246

Qy   81 TyrGlnProAspMetValAlaGlyTyrLeuSerLeuGlyValTyrSerAlaLeuValAlaSer 100
        :::||||||||||||||||||||||||||||||||||||||||||||||||
Db 247 GTTAAACCGGATAATGGTAGCTGGTCTTTCCTTAGGAGAATACTCAGCTTGGTAGCATCA 306

Qy 101 GlyAlaLeuAspPheGluAspAlaAlaAlaAlaAlaLeuAlaLeuAlaLeuAlaArgGlyAlaValTyrMetGlu 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 307 GGGGCACTCTCTTTGAAGATACCCTATCCTTAGTAGCTAAGAGAGGCCGCTTGATGGAG 366

Qy 121 GluAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
        ||||||||||||||||||||||||||||||||| |||||||| :::||
Db 367 GAAGCAGCACCACAAGGATCTGGAAAGATGGTTGCCGTTATGAATACAGACGTGCAAGTC 426

Qy 141 IleGluGlyAlaCysGlnLysAlaSerGluLeuGlyLeuValValThrProAlaAsnTyrAsn 160
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 427 ATCGAGGAGAGTCTGTCAAATAGCTGCTAAACATGGAGTGGTTGCTCCAGCAAACTATAT 486

Qy 161 ThrProAlaGlnIleValIleIleGluAlaAlaGlyGluValAspArgAlaValAlaValGluLeu 180
        :::|||||||||||||||||||||||||||| :::::
Db 487 ACTCCTAGTCAAATTGTAATTGGTGGTCAGACAGATGCTGTGAACGTGCAGTTGAACTT 546

Qy 181 LeuGlnGluAlaGlyAlaArgGlnLeuIleProLeuLysValSerGlyProPheHisThr 200
        |||:::|| ||||||||:::|| |||||||||||||||:::||
Db 547 TTAAGGAAGGGGAGTTAAGCGTTTAATCCCTTAAACGTGTCAGGTCCTTTCCACACT 606

Qy 201 AlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSer 220
        |||:::|| ::   :::   ::|||||||||||||||||||||||||
Db 607 GCTTTGTTGAACCAGCTAGCCGCTTGTTGGCTAAGAGTTGGAAAGATACAACTTCAGT 666

Qy 221 AspPheThrCysProLeuValGlyAsnThrGlnAlaAlaValMetGlnLysGluAspIle 240
        |||||||||||:::   :::|| :::||| :::
Db 667 GACTTCAAGATTCCTTGGTTGGTAATACCGGAGCTAATATTATGGAAAAAGACCGTATC 726

Qy 241 AlaGlnLeuThrArgGlnValLysGluProValArgPheGluSerIleGlyVal 260
        |||:::   :::|||||||||||||||||||||||||||||||
Db 727 CCAGAACTATTAGCCGTCAAGTCAAGTCATGGAGCCTGTTCGTTTTATGACAGTGTGCGACT 786

Db 783472 CATACAGCAATTTAAAACCAGCATCTGAAAAATTGGCTTTGGAGCTTGATGAAATTGAT 783531
         ||||||||||::|||||||| |||||||||||||||||||| ::::::
Qy   219 PheSerAspPheThrCysProLeuValGlyAsnThrGlyAlaAlaValMetGlnLysGlu 238
          ||||||||||::: |||||| :::||:::||
Db 783532 TTAGACCTTTGAATACCATTAATCCAAATACGAGTGTAAGTAATGGAAAAATGAT 783591

Qy   239 AspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheGluTyrGluSerIle 258
          ::::: ||||||||||||||||||||||||||||||||||||||||:::
Db 783592 GAAGTCAAAGACTTTGACGGTCAAGCTCAAGTCATGGAACCCAGTTCGTTTTTATGAATCGGTT 783651

Qy   259 GlyValMetGlnGluAlaGlyIleLeuSerAsnPheIleGluIleGlyProGlyLysValLeu 278
          |||||||:::   ::: |||||||||||:::||||||||
Db 783652 GAAACAATGCAAAAACTAGGGGCGACTCGCTTAATTGAAGTGGTCCTGGGAGAGTACTT 783711

Qy   279 SerGlyPheValLysIleAspPheGlyIleThrAlaHisAlaHisValGluAspGlnAla 298
          |||||||:::   :::   :::||| ::|||||||||:::||:::
Db 783712 TCAGGTTTCATTAAAAAAATTGATAAAAATGCAGAAATTGCTAATGTTGAAAATTTAGCT 783771

Qy   299 SerLeuValAlaLeuLeuGluLys 306
          |||::|||||||||:::
Db 783772 TCATTGAAGCTTGATTAATCAG 783795

RESULT 8
ABN68641
ID   ABN68641 standard; DNA; 942 BP.
XX
AC   ABN68641;
XX
DT   01-JUL-2002 (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 5195.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus pyogenes.
XX
PN   WO020234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-PSDB; ABP28010.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein.
XX
PS   Claim 7; Page 3863; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a

```
Alignment Scores:
Pred. No.:              7.99e-94        Length:          924
Score:                  996.50          Matches:         203
Percent Similarity:     79.08%          Conservative:     39
Best Local Similarity:  66.34%          Mismatches:       63
Query Match:            65.65%          Indels:            1
DB:                     24              Gaps:              1

US-09-308-397-2 (1-306) x ABN68640 (1-924)

Qy     1 MetThrLysThrAlaPheLeuPheAlaGlyGlnG-yAlaGlnTyrLeuGlyMetGlyArg  20
           :::||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db     1 ATGAATAAAGTTCATTTCATTTGCTGGTCAGGGTGCTCAAAGTGCTCAAAAGTTAGGAATGGGAATGGCTAGA  60

Qy    21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnAlaValLeu  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GATTTATATGAACATTTCCCGATTGTAAAGAAAACTTTGATAGGCAAGTCAAGTCATGTTTTA  120

Qy    41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyr  60
           |||||||||||||||||||||||||||||||||||::|::::::|||:::::||:::
Db   121 GGATATGATTTGAGAGAATTGATTGATACAGAAGACTTAAGCAAATTAAATCAGACGAGAAATAT  180

Qy    61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeu--LeuGlnGluLys  79
           ||||||||||||||||||||||||||||||||||||::|::|||  |::|:||::|||
Db   181 ACGCAACCAGCTATTCTAACGACTTCGACTGCTATTTATCGTTTATTCTTAAAGAAATT  240

Qy    80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla  99
           ::|||||||||||||||::||||||||||||||||||||||||||||||||||
Db   241 GAACTTAGACCAGATATGGTTGCAGGACTTTCTTTGGGAGAATACAGTGCTTTAGTTGCT  300

Qy   100 SerGlyValAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet  119
           ||||:|:||:|||::|||::||||:|:|||||||||||||||::|::||::|:||:::::
Db   301 TCAGGTGCTATAAGATTTGAAGATGCTGTAGCACTTGTTGCTAGACGAGGCCAACTAATG  360

Qy   120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu  139
           ||:||||||||||||||||::||||||||||||||||||||||||||||||||||||||
Db   361 GAAGCAGCAGCACCTGCTGGTAGTGCAAAATGGTAGCAGTTTAAATGCTGATCGTCAA  420

Qy   140 ValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyr  159
           :::|||||||||||::||||::|||||||||||||||||||:::::|||:::|||:::
Db   421 ATAATTGAAGATGCTTGCAGAAGGCTTCAGTTTGGTATTGTTAGTCCTGCAAATTAT  480

Qy   160 AsnThrProAlaGlnIleValIleLeuAlaGlyGluValAlaValAlaAspArgAlaValGlu  179
           ||||||||||||::|:||:|||||||||||||||:||||||||||||||::||::|||
Db   481 AATACTCCAAAGCAAATTGTTATAGGGGAGGAGTCAATCGCTGTAAATGCACTGTGAGAA  540

Qy   180 LeuLeuGlnGluAlaGlyLysArgLeuIleProLeuLysValSerGlyTyrProPheHis  199
           |||||::|||||:|:||||||::|||:|||:||||||::||||:||:||:|||||:|||
Db   541 GAATTAAAACAACAAGGAGTAAAACGTTTAATTCCTCTGAAATGTTTCAGGTCCTTTCAC  600

Qy   200 ThrAlaLeuGluProAlaSerGlnLysLeuLeuAlaGlnIleValSerPhe  219
           |||:||:|:|||||::||:::||||||::|||::|::|||:||::|:|::|
Db   601 ACTGCTTTATTGAAACCTGCAAGTCAAGTCAGAAATTGAGCGATGTTTTAGATAAAGTTCATTTT  660

Qy   220 SerAspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp  239
           ::||:::|::::|||::|:||||||::||||||||||||||::|||||::|:|||::|
Db   661 TCAGTTTCTGAAATCCCTGTATTGAAATACAGAGGCGCAGATAATGAAAAAGAGACGAT  720

Qy   240 IleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGly  259
           :::|||||::||:::|::||::|:||::|:|||:||::|:|::|:|||::|::|:|||
Db   721 ATCAAATCTTTATTGGCACGCCAAGTTAAGGAACCTGTACGCTTTGATGAGTCTATTGAA  780

Qy   260 ValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSer  279
           |||:||||||::|||:|::|::|||||:::|:|||||||||||::|::|:|:|:|:||:
Db   781 ACCATGAAAAAATGGAATGACACAGGTTGTTGAAATTGGTCCAGGGAAGGTACTCCTCA  840

Qy   280 GlyPheValLysLysGlnIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSer  299
           ||:||:||||:||:|||:|||||||:||::||::|:|||:::|:|:||:|::|||:|:|||
Db   841 GGTTCTTGAAAAAAATAGATAGCTCATTATCAGTACAATTCGTTGTTGAAGATAAAATTGGA  900

Qy   300 LeuValAlaLeuGlu  305
           |||||||||||||||
Db   901 TTTAAATAATCTTAAGGAA  918

Qy   261 MetGlnAlaGlyIleSerAsnPheGluIleGlyProGlyLysValSerGly  280
           :::   ||:||||||||||||||||||||||||||||||||||  ||:|||
Db   787 TTAGTAGAAAGTGGCATAACACAATTCATTGAGGTAGGACCAGGTAAGGTTTGACAGGT  846

Qy   281 PheValLysLysAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu  300
           |||||||||||||||:||||||||||||||||||||||||::|::||:||
Db   847 TTTGTGAAGAAAATTGATAAAAATTACTAGTGTTGAAAACATGGTAAGCTTA  906

Qy   301 ValAlaLeuGluLys  306
           :::::::|||||:|:
Db   907 CGTTTATTTTTAGATAGA  924

RESULT 9
ABN68640
ID   ABN68640 standard; DNA; 924 BP.
XX
AC   ABN68640;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 5193.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus agalactiae.
XX
PN   WO20234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GNO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI   Tetelin H;
XX
DR   WPI; 2002-352536/38.
DR   P-SDB; ABP28009.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 7; Page 3863; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassay, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 924 BP; 311 A; 132 C; 194 G; 287 T; 0 other;
```

RESULT 10
ABN71527
ID   ABN71527 standard; DNA; 2155561 BP.
XX
AC   ABN71527;
XX
DT   02-JUL-2002 (first entry)
XX
DE   Streptococcus polynucleotide SEQ ID NO 10967.
XX
KW   Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW   group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW   antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS   Streptococcus sp.
XX
PN   WO200234771-A2.
XX
PD   02-MAY-2002.
XX
PF   29-OCT-2001; 2001WO-GB04789.
XX
PR   27-OCT-2000; 2000GB-0026333.
PR   24-NOV-2000; 2000GB-0028727.
PR   07-MAR-2001; 2001GB-0005640.
XX
PA   (CHIR-) CHIRON SPA.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI   Tettelin H;
XX
DR   WPI; 2002-352536/38.
XX
PT   New Streptococcus protein for the treatment or prevention of infection
PT   or disease caused by Streptococcus bacteria, such as meningitis, and
PT   for detecting a compound that binds to the protein -
XX
PS   Claim 8; Page 4196-4488; 4525pp; English.
XX
CC   The invention relates to a protein (ABP25413-ABP30895) from group B
CC   streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC   (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC   the specification. The proteins have antibacterial and antiinflammatory
CC   activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC   antibodies that bind (I) are used in the manufacture of medicaments for
CC   the treatment or prevention of infection or disease caused by
CC   Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC   Nucleic acids encoding (I) are used to detect Streptococcus in a
CC   biological sample. (I) is used to determine whether a compound binds to
CC   (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC   used as a vaccine or diagnostic composition. The disease caused by
CC   Streptococcus that is prevented or treated may be meningitis. Nucleic
CC   acid encoding (I) may be used to recombinantly produce (I) and may be
CC   used in gene therapy. Antibodies to (I) are used for affinity
CC   chromatography, immunoassays, and distinguishing/identifying
CC   Streptococcus proteins.
XX
SQ   Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;

Alignment Scores:
Pred. No.:   2155561
Score: 2.92e-89   Length: 2155561
Percent Similarity: 79.08%   Matches: 203
Best Local Match: 66.34%   Conservative: 39
Query Match: 65.65%   Mismatches: 63
DB: Indels: 1
                                      Gaps: 1

US-09-308-397-2 (1-306) x ABN71527 (1-2155561)

Qy   1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyTyrMetGlyArg 20
           ||| || |||||| |||||||||||||||||||||||||||||||||
Db 330947 ATGAATAAAGTTTCATTTGCTGCTGCTCAAGTGCTCAAAAGTTAGGATGGCTAGA 331006

Qy   21 AspPheTyrAspGlnTyrProLeuValLysGluThrIleAspArgAlaSerGlnValLeu 40
           ::: ||| ::: |||
Db 331007 GATTATATGAAACTTTCCGATTGTAAAGAAACTTTTGATAAGGCAAGTCATGTTTTA 331066

Qy   41 GlyTyrAspLeuArgTyrLeuIleAspThrGluLysAspLysAsnGlnThrArgTyr 60
           ||| ::: :: :
Db 331067 GGATATGATTTGAGAGAATTGATGATAAAGACTTAGACAAATTAAATCAGACAGAAATAT 331126

Qy   61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleThrTyrArgLeu--LeuGlnGluLys 79
           ||| |||||||
Db 331127 ACGGAACCAGCTATTCTAACGACTTCGACTGCTATTTATCGTTTAATTCTTAAGGAAATT 331186

Qy   80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAla 99
           ::: |||||||||||| |||||||||||||||
Db 331187 GAACTTAGACCAGATATGGTTGCAGGACTTTCTTTGGGAGAATACAGTGCTTTAGTTGCT 331246

Qy   100 SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMet 119
           ||||||| |||||||||||||||
Db 331247 TCAGGTGCTATAAGATTTGAAGATGCTGTAGTACTTGTTGCTAGACGAGGCCAACTAATG 331306

Qy   120 GluGluAlaLeuAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu 139
           ||||||||| ::: :::
Db 331307 GAAGCAGCAGCACCTGCTGGTAGTGGCAAAATGGTAGCAGTTTTAAATGCTGATCGTCAA 331366

Qy   140 ValIleGluValGlyGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyr 159
           ::: ||||||| |||||
Db 331367 ATATTGAAGATGCTGTAAGAAGGCTTTCCAGTTTGGTATTGGTATTGGTATTGGT... 331426

Qy   160 AsnThrProAlaGlnIleIleAlaGlyValValAlaAlaAspArgAlaValAlaGlu 179
           ||||||| ::: :::
Db 331427 AATACTCCAAAGCAAATTGTTATAGGGGAGAGTCAATCGCTGTAAATCAGCTGAGAA 331486

Qy   180 LeuLeuGlnGluAlaGlyValAlaLysArgLeuIleProLeuLysValSerGlyProPheHis 199
           ||| ::: :::
Db 331487 GAATTAAAACACAAGGAGTAAAACGTTTAATTCCTGAAGTTTCAGGTCCTTTTCAC 331546

Qy   200 ThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGlnThrLeuGlnValSerPhe 219
           ||| |||
Db 331547 ACTGCTTTATTGAAACCTGCAAGTCAGAAATTGAGCGATGTTTAGATAAAGTTCATTTT 331606

Qy   220 SerAspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp 239
           ||| ::: :: |
Db 331607 TCAGTTTCTGAAATCCCTGTAATTGGAAATACAGAGGCGCAGATAATGAAAAAAGACGAT 331666

Qy   240 IleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGly 259
           ||| |||
Db 331667 ATCAAATCTTTATTGGCACGCCAAGTTATGGAACCTGTACGCTTGATGAGTCTATTGAA 331726

Qy   260 ValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSer 279
           |||||| ::: :::
Db 331727 ACCATGAAAAAATGGAATAACACACAGGTTGTTGAAATTGGTCCAGGAAGTACTCTCA 331786

Qy   280 GlyPheValLysIleLysAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSer 299
           ||||||| ||| :::
Db 331787 GGTTTCTGAAAAAATAGATAGATGCTCATTATCAGTAGATAAAATTGGA 331846

Qy   300 LeuValAlaLeuLeuGlu 305
           |||
Db 331847 TTTAATAATCTTAAGGAA 331864

RESULT 11
AAS53186
ID   AAS53186 standard; DNA; 942 BP.
XX
AC   AAS53186;
XX
DT   13-FEB-2002 (first entry)
XX
DE   Enterococcus faecalis DNA for cellular proliferation protein #614.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;

KW   antibiotic; antibacterial; drug design.
XX
OS   Enterococcus faecalis.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI   Yamamoto RT, Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU35327.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27, SeqID No 6823; 51pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 942 BP; 300 A; 166 C; 234 G; 242 T; 0 other;

Alignment Scores:
Pred. No.:   942
Score: 1.2e-87   Matches: 937.00   187
Percent Similarity: 75.17%   Conservative: 40
Best Local Similarity: 61.92% Mismatches: 75
Query Match: 61.73%   Indels: 0
DB: 23   Gaps: 0

US-09-308-397-2 (1-306) x AAS53186 (1-942)

Qy   3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyTyrArgAspPhe 22
                                          :::
Db   16 AAAACAGCGATTTATTTAGTGGACAAGGAGCCCAGTATCAAGGGATGGGTGAAGAATTA 75

Qy   23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerArgGlnValLeuGlyTyr 42
Db   76 TATCCAGAAGAGCGATTGTTCGGGAAACTTTCGATGAAGCAAGTCATATCTTAGGTTAT 135

Qy   43 AspLeuArgTyrLeuIleAspThrGluGluAspLeuAspLeuAsnLeuAsnGlnThrArgThrGln 62
                                          :::::::
Db   136 GAGATGGCAGAACTTGTTTACTGAAAATGAAGTTTAAATGAAACAGAATATACGCAA 195

Db   63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGluGlnLysGlyTyrGln 82
Qy   196 CCTGCTATTTTAACAGTCAGTCAGTGTCGCATTTTACCGTCTTTTGCAACAAAAAGGACTAACG 255

Db   83 ProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAlaSerGlyAla 102
Qy   256 CCTGATGTCGTAGGCGGGTTTAAGCTTAGGGGAATATAGCGCTTTGGTTGCCAGCGGGGGCT 315

Qy   103 LeuAspPheGluArgGlyAlaLeuAlaLeuAlaLysArgGlyAlaTyrMetGluGluAla 122
Db   316 TTGCGCTTTCAGAAGCAGTGGCCCTTGGTCCAGACATGACAGAAGCA 375

Qy   123 AlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGlu 142
Db   376 GCACCACAAGGAACTGGCAAAATGGTTGCTGTCATGAATGCTGAGCGTGAAGTAATTGAG 435

Qy   143 GluAlaCysGlnLysAlaSerGluGlyGlyValValThrProAlaAsnTyrAsnThrPro 162
Db   436 AAAGCCTGCCAAGAAGCCAGTGCTTTCGGAATTGTGGCTCCAGCAAATTATAATACACCA 495

Qy   163 AlaGlnIleValIleLeuAlaGlyValGluValAlaAlaValAspArgAlaValGluLeuGln 182
Db   496 CAACAAATCTGATTGGTGGTGAGGTTGCTGCTGTTGATCAAGCGATGACACTTCTCAA 555

Qy   183 GluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeu 202
Db   556 GAAGCTGGTGTGAGAAGCGAATGATTCCGTTAAATGTGAGTGCGCCCTTTCCATACGGCGTG 615

Qy   203 LeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnLysValSerPheSerArgPhe 222
Db   616 TTACACCAGCATCAAAAAAATTGGCTCAGGATTTAGCAAAATTGAACTTTCAAACGATG 675

Qy   223 ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAlaGln 242
Db   676 CAAATTCCTGTCATTAGTAATACGACTGCCGAAAATTATGCCCAAGAGGCAATTCAGCG 735

Qy   243 LeuLeuThrArgGlnValGlyProValArgPheTyrGluSerIleGlyValMetGln 262
Db   736 TTATTGGAAAAGCAAGTCATGTCTGCGGTACGTTTTGAAGACAGTATCGAAACGATGAAG 795

Qy   263 GluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheVal 282
Db   796 GCTATGAACGTGGAACGATGATTGAAGTTGGTCCAGGGAAAACATTAACTGGTTTTGTT 855

Qy   283 LysLysIleAspGlnThrArgAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAla 302
Db   856 AAAAAAATTGACAAACAATTGAAATGCACCGTGTGGAAGATGTTGCAACATTAACAGAA 915

Qy   303 LeuLeu 304
Db   916 ACGTTA 921

RESULT 12
AAX13514
ID   AAX13514 standard; DNA; 3656 BP.
XX
AC   AAX13514;
XX
DT   19-MAR-1999 (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:577.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX

```
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC, Dillon PJ, Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 1834-1836; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis infection in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 3656 BP; 1210 A; 627 C; 837 G; 979 T; 3 other;
```

Alignment Scores:

```
Pred. No.:
Score:                  4.35e-85          Length:        3656
Percent Similarity:     75.25%            Matches:       186
Best Local Similarity:  62.21%            Conservative:  39
Query Match:            60.61%            Mismatches:    73
DB:                     20                Indels:        1
                                          Gaps:          0
```

US-09-308-397-2 (1-306) x AAX13514 (1-3656)

```
Qy     3 LysThrAlaPheLeuPheAlaAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
           |||  ||| |||||||||    |:::
Db  2753 AAAACAGGATTTATTTGTGACAAGGAGCCCAGTATCAAGGGGATGGGTGAAGAATTA 2812

Qy    23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerArgAlaSerGlnValLeuGlyLyTyr 42
           ||||||||||||||| |||:::|||||||
Db  2813 TATCACCAAGAGAGCGATTGTTCGGGAAACTTTCGATGAAGCAAGTCATATCTTAGGTTAT 2872

Qy    43 AspLeuArgTyrLeuIleAspThrGluLeuAspLysLeuAsnGlnThrArgTyrThrGln 62
           |||  :::::::|||
Db  2873 GAGATGGCAGACTTTGTTTTACTGAAAATGAAACAGTTTAAAATGAAACAGAATATACGCAA 2932

Qy    63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnLeuGlyLysGlyGlyTyrTyrGln 82
           ||||||  |||
Db  2933 CCTGCTATTTTAACAGTCCAGTGCGCATTTTACCGGTCTTTGCAACAAAAGGACTAACG 2992

Qy    83 ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyAlaAla 102
           ||||||||||| :::|||||||||||||||||
Db  2993 CCTGATGTCGTAGCGGGTTTAAGCTTAGGGGAATACAGTGCTTTGGTTGCCAGCGGGGCT 3052

Qy   103 LeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAla 122
           |||| :::|||||||
Db  3053 TTGGCGTTTTCAGAAGCAGTGGCCTTGGTCCAAAAGCGCGGTGCAGTACATGACAGAGCA 3112

Qy   123 AlaProAlaAspSerGlyLyMetValAlaValAlaLeuAsnThrProValGluValIleGlu 142
           :::::::|||||||
Db  3113 GCACCACAAGGAACTGGCAAAATGGTTGCTGTTGCAGTAATGAATGCTGAGCGTGAGTAATTGAG 3172
```

```
Qy   143 GluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrPro 162
           :::|||||||||||||||||  :::
Db  3173 AAAGCCTGCCAAGAGCCAGTGCTTCGGAATTGGCTCCAGCAAATTATAATACCACCA 3232

Qy   163 AlaGlnIleValIleAlaGlyIleValGlnValAlaValAlaAspArgAlaValGluLeuGln 182
           ||||||||||||||| |||:::|||||||
Db  3233 CAACAAATCGTGATTGCTGGGATTGTCGTGTGATCAAGCAGCACTTCTCAAA 3292

Qy   183 GluAlaGlyAlaLysArgLeuIleProLeuLysSerGlyProPheHisThrAlaLeu 202
           |||  :::|||
Db  3293 GAAGCTGGTGTGAAGCGAATGATCCGTTAAAATGAGTGAGGGCCCTTCCATACGGCGCTG 3352

Qy   203 LeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPhe 222
           |||||:::|||||||
Db  3353 TTACAACCAGCATCAAAAAAATGGCTCAGGATTTAGCAAAATTGAACTTTCAAACGATG 3412

Qy   223 ThrCysProLeuValGlyAsnThrGlyAlaAlaValMetGlnLysGluAspIleAlaGln 242
           |||||  :::::::|||
Db  3413 CAAATTCCTGTCATTAGTAATACGACTGCCGAAATTATGCCCCAGAGGCAATTCAAGCG 3472

Qy   243 LeuLeuThrArgGlnValValGluProValArgPheTyrGluSerIleGlyValMetGln 262
           ||||||  |||
Db  3473 TTATTGGAAAAGCAAGTCATGTCTGCGTACGTTTTGAAGACAGTATCGAAACGATGAAG 3532

Qy   263 GluAlaGlyIleSerAsnPheIleGluIle-GlyProGlyLysValLeuSerGlyPheVal 282
           :::|||||||
Db  3533 GCTATGAACGTAGGAACGATGATTGAAGTTNGGTCCAGGGRAAACATTAACYGGTTTGT 3592

Qy   282 lLysLysIleAspThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu 300
           :::|||||||
Db  3593 TAAAAAAATTGACCAAACAATTGAAATGACCGTGGAAGATGTTGCACATTA 3647
```

```
RESULT 13
AAV37241
ID   AAV37241 standard; DNA; 547 BP.
XX
AC   AAV37241;
XX
DT   24-SEP-1998 (first entry)
XX
DE   C-terminal coding region encoding a FabD polypeptide embodiment.
XX
KW   FabD protein; malonyl-CoA:ACP family; diagnosis; infection; vaccine;
KW   screen; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO982133-A1.
XX
PD   28-MAY-1998.
XX
PF   14-NOV-1997;   97WO-US20992.
XX
PR   18-NOV-1996;   96US-0031160.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Gentry DR, Lonsdale JT, Payne DJ, Pearson SC, Van Aller G;
XX
DR   WPI; 1998-312173/27.
DR   P-PSDB; AAW60858.
XX
PT   New isolated Streptococcus pneumoniae FabD gene - used to develop
PT   products for the diagnosis, prevention and treatment of bacterial
PT   diseases, particularly S. pneumoniae infection.
XX
PS   Disclosure; Pages 6-7; 45pp; English.
XX
CC   The present sequence represents the C-terminal coding region of FabD
CC   embodiment (coding begins at nucloetide 2) of Streptococcus pneumoniae
CC   0100093 (NCIMB 40800). The novel FabD polypeptides are related to other
CC   proteins of the malonyl-CoA:ACP family. The products can be used for the
```

```
CC   diagnosis of Streptococcus pneumoniae infections. Vectors containing
CC   the FabD DNA sequence can be administered directly to a mammal to
CC   produce the FabD peptide to provoke an antibody/T-cell response in
CC   order to prevent a disease. The peptide can be used to screen for
CC   compounds which modulate its activity.
XX
SQ   Sequence 547 BP; 154 A; 115 C; 132 G; 146 T; 0 other;

Alignment Scores:
Pred. No.:              2.67e-80        Length:        547
Score:                  863.00         Matches:       176
Percent Similarity:     97.25%         Conservative:  1
Best Local Similarity:  96.70%         Mismatches:    5
Query Match:            56.85%         Indels:        0
DB:                                    Gaps:          0

US-09-308-397-2 (1-306) x AAV37241 (1-547)

Qy 125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
Db   2 GCTGAACTTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTCATTGAAGAAGCC  61

Qy 145 CysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGln 164
Db  62 TGTCAAAAGCTTCTGGAACTTGGAGTGGTTACTCCAGCCAACTATAACACCTGCACAA  121

Qy 165 IleValIleAlaGlyGluValValAlaValAlaAlaValAspArgAlaValAlaGluGluAla 184
Db 122 ATCGTCATTGCTGGAGAAGTGGTTGCAGTTGATCGAGCGGTTGAACTTTGCAAGAAGCA 181

Qy 185 GlyAlaLysArgLeuIleProLeuLysSerGlyProPheHisThrAlaLeuLeuGlu 204
Db 182 GGTGCCAAACGCTTGATTCCTCTTAAGAGTGGTCAGGTCCCTTTCACACGGCTCTCCTGAG 241

Qy 205 ProAlaSerGlnLysLeuLeuAlaGluThrPheSerAspPheSerThrCys 224
Db 242 CCTGCTAGCCAGAAACTAGCTGAAACTTCTAGCTCAGGTAAGTTTTCAGATTTTACTTGT 301

Qy 225 ProLeuValGlyAsnThrGluIleMetGlnLysGluAspIleAlaGlnLeuLeu 244
Db 302 CCCCTAGTCGGCATACAGAAGGCTGCTGATGCAAAAAGAGGACATTGCCAGCCTCTTG 361

Qy 245 ThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGlnValAla 264
Db 362 ACGCGTCAGGTCAAGGAACCCGTTCGTTTCTATGAAAGTATTGGGGTCATGCAAGAGCA 421

Qy 265 GlyIleSerAsnPheIleGlyIleLeuGlyProGlyLeuValLeuSerGlyLysPheLys 284
Db 422 GGCATAAGCAACTTCATTGGAATTGGACCGGGGAAAGTCTTGTCAGGTTTTGTTAAAAAA 481

Qy 285 IleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuLysValAlaLeuLeu 304
Db 482 ATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCCAGTTTAGTAGCACTTTTA 541

Qy 305 GluLys 306
Db 542 GAAAAA 547
```

```
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             586..1146
FT                   /*tag= a
XX
PN   WO9743303-A1.
XX
PD   20-NOV-1997.
XX
PF   14-MAY-1997;   97WO-US07950.
XX
PR   14-MAY-1996;   96US-0017670.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1998-008793/01.
DR   P-PSDB; AAW38474.
XX
PT   Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT   diagnosing anti-microbial agents for treatment of bacterial
PT   infections
XX
PS   Claim 4; Page 90; 483pp; English.
XX
CC   This sequence encodes a Streptococcus pneumoniae protein that (based on
CC   homology with a Cuphea lanceolata protein) is a 3-oxoacyl reductase
CC   precursor protein, and represents a DNA sequence of the invention.
CC   The DNA sequences were isolated from Streptococcus pneumoniae strain
CC   0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC   invention can be used to identify compounds which interact with and
CC   inhibit or activate the activity of the proteins. Antagonists can be
CC   used to treat diseases caused by S. pneumoniae proteins, through genetic
CC   immunisation. They can also be used to induce an immunological response
CC   in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC   of the encoding nucleic acids in a vector adequate to produce antibody
CC   and/or T cell immune responses to protect the animal from disease. The
CC   proteins can also be used to identify antimicrobial compounds which are
CC   capable of inhibiting their bioactivity. In particular the proteins of
CC   the invention can be used to prevent adhesion of bacteria to mammalian
CC   extracellular matrix proteins on in-dwelling devices or in wounds, to
CC   block protein-mediated mammalian cell invasion, and to block the normal
CC   progression of pathogenesis in infections initiated other than by the
CC   implantation of in-dwelling devices or other surgical techniques.
XX
SQ   Sequence 1196 BP; 355 A; 223 C; 286 G; 332 T; 0 other;

Alignment Scores:
Pred. No.:              2.54e-79        Length:        1196
Score:                  858.00         Matches:       180
Percent Similarity:     98.37%         Conservative:  1
Best Local Similarity:  97.83%         Mismatches:    2
Query Match:            56.52%         Indels:        2
DB:                                    Gaps:          0

US-09-308-397-2 (1-306) x AAT98542 (1-1196)

Qy 125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
Db   2 GCTGAACTTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTTCTCAATACGCCAGTAGAGGTCATTGAAGAAGCC  61

Qy 145 CysGlnLysAlaSer-GluLeuGlyValGlyValValThrProAlaAsnTyrAsnThrProAlaGl 164
Db  62 TGTCAAAAGCTTCTGGAACTTGGAGTGGTGGTTACTCCAGCCAACTATAACACCTGCACA 121

Qy 164 nIleValIleAlaGlyGluValValAlaValAlaValAlaAspArgAlaValAlaGluLeuGlnGluAl 184
Db 122 AATCGTCATTGCTGGAGAAGTGGTTGCAGTTGCAGTTGATGAGCGGTTGAACTTTGCAAGAAGC 181
```

```
RESULT 14
AAT98542
ID   AAT98542 standard; DNA; 1196 BP.
XX
AC   AAT98542;
XX
DT   06-NOV-1998  (first entry)
XX
DE   DNA encoding a S. pneumoniae 3-oxoacyl reductase precursor protein.
XX
KW   Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW   immunological response; inoculation; antibody production; inhibitor;
KW   T cell immune response; antimicrobial compound; bacterial adhesion;
KW   extracellular matrix protein; protein-mediated cell invasion; wound;
KW   pathogenesis; ss.
```

```
CC      be used to treat an individual having need to inhibit a bacterial
CC      protein. Vectors expressing the proteins can be used to induce a
CC      protective immune response in mammals.
XX
SQ      Sequence 1199 BP; 334 A; 287 C; 223 G; 355 T; 0 other;

Alignment Scores:
Pred. No.:                    2.55e-79        Length:          1199
Score:                        858.00          Matches:          180
Percent Similarity:           98.37%          Conservative:       1
Best Local Similarity:        97.83%          Mismatches:         1
Query Match:                  56.52%          Indels:             2
DB:                           18              Gaps:               0

US-09-308-397-2 (1-306) x AAX30801 (1-1199)

Qy     125 AlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAla 144
            ||:|| ::::          ||||:|:||| |||:|:||||: |||||||||||||||||
Db    1198 GCTGAACTTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTCATTGAAGAAGC 1139

Qy     145 CysGlnLysAlaSer-GluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGl 164
            |||||||||||||| ||||||||||||||| ||||:|||  ||||||||||||||||||
Db    1138 TGTCAAAAGTCTCTGGAACTTGGAGTGGTTACTCCGCCAACTATAACACACCTGCACA 1079

Qy     164 nIleValIleAlaGlyGluValValAlaValAlaSerArgAlaValGluLeuGlnGluLeuGluI 184
            |||||||||||||||||||||||:||||||||||||| |||||||||||||||||||||
Db    1078 AATCGTCATTGCTGGAGAAGTGGTTGCAGTTGCATCGAGCGGGTTGAACTTTGCAAGAAGC 1019

Qy     184 aGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGl 204
            ||||||||||||||||||| ||||||||||||||| ||||||| ||||||||||||||||
Db    1018 AGGTGCCAAACGTTGATTCCCTTAAGGTGTCAGGTCCCTTTCACACCGCTCTCCTTGA 959

Qy     204 uProAlaSerGlnLysLeuAlaGluThrLeuGlnIleValSerPheSerAspPheThrCy 224
            |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     958 GCCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTTG 899

Qy     224 sProLeuValGlyAsnThrGluAlaAlaAlaValMetGlnLysGluAspIleAlaGlnLeuLe 244
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     898 TCCCCTAGTGGCAATACAGAAGCTGCTGCTGTGATGCAAAAAGAGGACATTGCTCAGGTCT 839

Qy     244 uThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGlnGluIleGluI 264
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     838 GACGCGTCAGGTCAAGAACCCGTTCGTTTCTATGAAAGTATTGGGGTCATGCAAGAAGC 779

Qy     264 aGlyIleSerAsnPheIle-GluIleGlyProGlyLysValLeuSerGlyValPheValLysL 284
            ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     778 AGGCATAAGCAACTTATTCGAGATTGGACCGGGGAAAGTCTTGTCAGGTTTGTTTAAAA 719

Qy     284 ysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuL 304
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     718 AAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTTT 659

Qy     304 euGluLys 306
            ||||||||
Db     658 TAGAAAAA 651

Search completed: June 11, 2003, 20:37:37
Job time : 1188 secs
```

```
Qy     184 aGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGl 204
            ||||||||||||||||||| |||||||||||||||| ||||||| |||||||||||||||
Db     182 AGGTGCCAAACGTTGATTCCCTTAAGGTGTCAGGTCCCTTTCACACCGCTCTCCTTGA 241

Qy     204 uProAlaSerGlnLysLeuAlaGluThrLeuGlnIleValSerPheSerAspPheThrCy 224
            |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db     242 GCCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTTG 301

Qy     224 sProLeuValGlyAsnThrGluAlaAlaAlaValMetGlnLysGluAspIleAlaGlnLeuLe 244
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 TCCCCTAGTGGCAATACAGAAGCTGCTGCTGTGATGCAAAAAGAGGACATTGCTCAGCTCTT 361

Qy     244 uThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGlnGluAl 264
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 GACGCGTCAGGTCAAGAACCCGTTCGTTTCTATGAAAGTATTGGGGTCATGCAAGAAGC 421

Qy     264 aGlyIleSerAsnPheIle-GluIleGlyProGlyLysValLeuSerGlyValPheValLysL 284
            |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db     422 AGGCATAAGCAACTTTATTCGAGATTGGACCGGGGAAAGTCTTGTCAGGTTTGTTTAAAA 481

Qy     284 ysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuL 304
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 AAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCGAGTTTAGTAGCACTTT 541

Qy     304 euGluLys 306
            ||||||||
Db     542 TAGAAAAA 549

RESULT 15
AAX30801/c
ID   AAX30801 standard; DNA; 1199 BP.
XX
AC   AAX30801;
XX
DT   20-MAY-1999  (first entry)
XX
DE   Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:78.
XX
KW   Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW   streptococcal infection; pneumococcal; ss.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9737026-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;  97WO-US05306.
XX
PR   22-AUG-1996;  96US-0025788.
PR   02-APR-1996;  96US-0014690.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1997-503111/46.
DR   P-PSDB; AAY11212.
XX
PT   Nucleic acids encoding pneumococcal polypeptide(s) - useful in
PT   vaccines, drug screening, etc
XX
PS   Claim 5; Page 105-106; 354pp; English.
XX
CC   AAX30724 to AAX30946 represent genomic DNA sequences isolated from
CC   Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
CC   encode the novel proteins given in AAY11114 to AAY11367. The proteins,
CC   isolated from Streptococcus pneumoniae, can be used in vaccines against
CC   streptococcal infections and in assays for identifying compounds that
CC   inhibit or activate the activity of the proteins. The antagonists can
```

OM protein - nucleic search, using frame_plus_p2n model

Copyright   GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: June 11, 2003, 20:12:23 ; Search time 68 Seconds
                                     (without alignments)
                                     1380.045 Million cell updates/sec

Title: US-09-308-397-2
Perfect score: 1518
Sequence: 1 MTKTAFLFAGQGAQYLGMGR.........QTAHLAHVEDQASLVALLEK 306

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop   6.0 , Delext   7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09308397/runat_06062003_112349_28904/app_query.fasta_1.455
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US09308397 @CGN_1_36 @runat_06062003_112349_28904 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents_NA:*
              1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
              2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
              3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
              4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
              5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
              6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
c 1	1518	100.0	19702	4	US-08-961-527-7	Sequence 7, Appli
2	858	56.5	1196	4	US-08-858-207A-8	Sequence 8, Appli
c 3	619	40.8	9686	4	US-09-221-017B-1003	Sequence 1003, Ap
4	547.5	36.1	927	1	US-08-789-609A-1	Sequence 1, Appli
5	547.5	36.1	927	1	US-09-108-517-1	Sequence 1, Appli
6	515.5	34.0	936	4	US-09-134-001C-1987	Sequence 1987, Ap
7	414.5	27.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
8	414.5	27.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
9	376.5	24.8	50937	4	US-09-428-517-1	Sequence 1, Appli
10	374	24.6	68750	3	US-09-335-409-1	Sequence 1, Appli
11	374	24.6	68750	4	US-09-568-102-1	Sequence 1, Appli
12	374	24.6	68750	4	US-09-567-969-1	Sequence 1, Appli
13	374	24.6	68750	4	US-09-568-480-1	Sequence 1, Appli
14	374	24.6	68750	4	US-09-568-486-1	Sequence 1, Appli
15	374	24.6	68750	4	US-09-568-472-1	Sequence 1, Appli
16	374	24.6	68750	4	US-09-567-899-1	Sequence 1, Appli
17	372.5	24.5	15872	4	US-09-105-537-1	Sequence 1, Appli
c 18	372	24.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 19	372	24.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
20	369	24.3	71989	2	US-09-443-501A-2	Sequence 2, Appli
21	367.5	24.2	43280	2	US-08-804-227C-1	Sequence 1, Appli
22	349.5	23.0	33529	4	US-09-144-085-3	Sequence 3, Appli
23	348.5	23.0	8268	1	US-08-375-709-10	Sequence 10, Appl
24	348.5	23.0	8268	1	US-08-752-929-10	Sequence 10, Appl
25	348.5	23.0	37895	1	US-08-375-709-1	Sequence 1, Appli
26	348.5	23.0	37895	1	US-08-752-929-1	Sequence 1, Appli
27	348.5	23.0	37895	1	US-09-090-793-1	Sequence 1, Appli
28	344.5	22.7	80161	3	US-09-036-987A-1	Sequence 1, Appli
29	344.5	22.7	80161	4	US-09-370-700-1	Sequence 1, Appli
30	343.5	22.6	19227	3	US-09-090-793-13	Sequence 13, Appl
31	343.5	22.6	40138	3	US-09-090-793-12	Sequence 12, Appl
32	334	22.0	1030	3	US-08-858-003-2	Sequence 2, Appli
33	334	22.0	1030	3	US-09-078-166-2	Sequence 2, Appli
34	334	22.0	1030	4	US-08-997-467-2	Sequence 2, Appli
35	329.5	21.7	49377	2	US-08-764-233A-1	Sequence 1, Appli
36	324	21.3	13842	4	US-09-105-537-30	Sequence 30, Appl
37	324	21.3	36778	5	US-09-105-537-5	Sequence 5, Appli
38	324	21.3	38506	3	US-09-320-878-19	Sequence 19, Appl
39	316.5	20.8	28958	1	US-08-258-261B-6	Sequence 6, Appli
40	316.5	20.8	28958	1	US-08-456-837-6	Sequence 6, Appli
41	316.5	20.8	28958	1	US-08-457-342-6	Sequence 6, Appli
42	316.5	20.8	28958	1	US-08-457-646A-6	Sequence 6, Appli
43	316.5	20.8	28958	1	US-08-458-076A-6	Sequence 6, Appli
44	316.5	20.8	28958	1	US-08-764-233A-4	Sequence 4, Appli
45	316.5	20.8	28958	1	US-08-457-335A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-961-527-7/c
; Sequence 7, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
;   APPLICANT: Charles Kunsch
;   TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
;   NUMBER OF SEQUENCES: 391
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/961,527
;   FILING DATE:
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Brookes, A. Anders
;   REGISTRATION NUMBER: 36,373
;   REFERENCE/DOCKET NUMBER: PB340P1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:

```
;    LENGTH: 19702 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: double
;    TOPOLOGY: linear
US-08-961-527-7

Alignment Scores:
Pred. No.:                           19702
Score:                   1518.00      Length:             19702
Percent Similarity:      100.00%      Matches:              306
Best Local Similarity:   100.00%      Conservative:           0
Query Match:             100.00%      Mismatches:             0
DB:                                   Indels:                 0
                                      Gaps:                   4
```

```
US-09-308-397-2 (1-306) x US-08-961-527-7 (1-19702)

Qy       1 MetThrLysThrAlaPheLeuPheAlaGlyLnGlnAlaGlnTyrLeuGlyMetGlyArg  20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17149 ATGACTAAAACAGCCTTTTTATTTGCTGGTCAAGGTGCCCAGTATCTAGGGATGGGACGG 17090

Qy      21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17089 GATTTCTATGATCAGTATCCGATTGTTAAAGAAACGATTGATCGAGCGAGCCAGTGCTC  17030

Qy      41 GlyTyrTyrAspLeuArgTyrLeuLeuAspThrGluAspLysLeuAsnGlnThrArgTyr  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17029 GGTTATATGATTTGCGTTATCCATCAGATAACGGAAGAAGACAAACTCAATACAGACCCGCCTAT 16970

Qy      61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGly  80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16969 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC 16910

Qy      81 TyrGlnProAspMetValAlaGlyLeuValLeuSerLeuGlyValGlyTyrSerAlaLeuValAlaSer 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16909 TATCAGCCTGATATGGTTGCTGGTTTGGTCCTTGCCTTGGTGGCAAGT 16850

Qy     101 GlyAlaLeuAspPheGluAspAlaAspSerGlyLysAlaLeuValAlaLeuLysAlaArgGlyAlaTyrMetGlu 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16849 GGCGCCCTTGGATTTGAAGATGCGGTTGCCCTGGTAGCTAAGGCCTATATGGAA 16790

Qy     121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGluVal 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16789 GAAGCCGCTCCTGCTGACTCGGCAAGATGTAGCAGTTCTCAATACGCCAGTAGAGGTC 16730

Qy     141 IleGluGluAlaCysGlnLnLysGlnAlaSerGlnAlaSerGlyLysValValValProAlaAlaAsnTyrAsn 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16729 ATTGAAGAAGCCTGTCAAAAAGCTTCAAGCGCTGCTGGTGGTTACTCCAGCCAACTATAAC 16670

Qy     161 ThrProAlaGlnIleValIleGluAlaGlyGluValAlaValAlaAspArgAlaValGluLeu 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16669 ACACCTGCACAAATCGTCATTGCTGCAGGTGGTTGCAGTTGCAGCGGTTGAACTT 16610

Qy     181 LeuGlnGluAlaGlyAlaLysAlaGlyArgLeuIleProLeuLysGlyValSerGlyValProPheHisThr 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16609 TTGCAAGAAGCAGGTGCCAAACGCTTGAGTTCCTCTTAAGGTGTCAGGTCCCTTTCACACC 16550

Qy     201 AlaLeuGluProAlaSerGlnLnLysGlnLeuAlaGlyThrLeuAlaGlnValSerPheSer 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16549 GCTCTCCTTGAACCTGCTAGCCAGAAACTAGCTGAAACTCTGACATCTAGCTCAGGTAAGTTTTCA 16490

Qy     221 AspPheTyrCysProLeuValGlyValAsnThrGlyAlaAlaAlaValMetGlnLysGluAspGluAspIle 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16489 GATTTTTACTTGTCCCCTAGTCGATAACAGAAGAGTGCTGTGATGCAAAAGAGGACATT 16430

Qy     241 AlaGlnLeuGlnThrArgGlnLnValLysGluGluProValGlyProValArgPheTyrGluSerIleGlyLyrVal 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16429 GCTCAGCTCTGACCGGTCAGGTCAAGGACACCCGTTGTTTCGTTCTTCATGAAAGTATTGGGGTC 16370

Qy     261 MetGlnGluAlaGlyIleLeuSerAsnPheIleGluIleGlyProGlyLysGluGluGluIleGluGluGluGlyGluGlyAlaLysValLeuSerGly 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16369 ATGCAAGAAGCAGGCATAAGCAACTTATCGAGATTGGACCGGGGAAAGTCTTGTCAGGT 16310

Qy     281 PheValLysLysIleAspGlnThrArgAlaHisLeuAlaHisValAlaHisLeuAlaAspGlnAlaSerLeu 300
```

```
US-08-858-207A-8

Alignment Scores:
Pred. No.:           5.88e-91              Length:            1196
Score:               858.00               Matches:            180
Percent Similarity:  98.37%               Conservative:         1
Best Local Similarity: 97.83%             Mismatches:           2
Query Match:         56.52%               Indels:               0
DB:                                       Gaps:                 4
```

```
US-09-308-397-2 (1-306) x US-08-858-207A-8 (1-1196)

Qy     125 AlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGluValIleGluGluAla 144
           |||||||||  |:||
Db       2 GCTGAACTGGCAAGATGGTAGCAGTTCTCAATACGCCAGTAGAGGTCATTGAAGAAGCC  61

Qy     145 CysGlnLysAlaSer-GluLeuGlyValValValThrProAlaAsnTyrAsnThrProAlaGlI 164
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TGTCAAAAAGCTTCTGGAACTTGGAGTGGTTACTCCAGCCAACTATAACACACCTGCACA 121

Qy     164 nIleValIleAlaGlyGluValAlaValAlaAspArgAlaValGluLeuLeuGlnGluAl 184
```

```
Db  122 AATCGTCATTGCTGCAGTGAGAAGTGGTTGCAGTTGATCGAGCGGTTGAACTTTTGCAAGAAGC 181
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  184 aGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGl 204
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 AGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCAGGTCCCTTTCACACCGCTCTCCTTGA 241
Qy  204 uProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCy 224
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  242 GCCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTCAGATTTTACTTG 301
Qy  224 sProLeuValGlyAsnThrGluAlaMetGlnLysGluAspIleAlaGlnLeuLe 244
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  302 TCCCCTAGTGCCAATACAGAAGGCTGCTGTGATCGAAAAAGAGGACAATTGCTCAGCTCTT 361
Qy  244 uThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGlnGluAl 264
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  362 GACGCGTCAGGTCAAGGAACCCGTTCGTTTCATGAAAGTATTGGGGTCATGCAAGAAGC 421
Qy  264 aGlyIleSerAsnPheIle-GluIleGlyProGlyLyAlaValLeuSerGlyValPheValLysL 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  422 AGGCATAAGCAAACTTTATTCGAGATTGGACCGGGAAAAGTCTTGTCAGGTTTTGTTAAAA 481
Qy  284 ysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeuValAlaLeuL 304
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  482 AAATTGATCAAACTGCTCACTTAGCTCATGTGGAAGATCAAGCCAGTTTAGTAGCACTTT 541
Qy  304 euGluLys 306
        ||||||||
Db  542 TAGAAAAA 549
```

```
RESULT 3
US-09-221-017B-1003/c
; Sequence 1003, Application US/09221017B
; Patent No. 6444799
GENERAL INFORMATION:
    APPLICANT:  Ross, Bruce C.
    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
    NUMBER OF SEQUENCES:  1120
    CORRESPONDENCE ADDRESS:
        ADDRESSEE:  MORRISON & FOERSTER
        STREET:  755 PAGE MILL ROAD
        CITY:  Palo Alto
        STATE:  CA
        COUNTRY:  USA
        ZIP:  94304-1018
    COMPUTER READABLE FORM:
        MEDIUM TYPE:  Diskette
        COMPUTER:  IBM Compatible
        OPERATING SYSTEM:  Windows
        SOFTWARE:  FastSEQ for Windows Version 2.0b
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER:  US/09/221.017B
        FILING DATE:  23-DEC-1998
        CLASSIFICATION:
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  PP1182
        FILING DATE:  31-DEC-1997
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  PP1546
        FILING DATE:  30-JAN-1998
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  PP2911
        FILING DATE:  09-APR-1998
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER:  PCT/AU98/01023
        FILING DATE:  10-DEC-1998
    ATTORNEY/AGENT INFORMATION:
        NAME:  Monroy, Gladys H
        REGISTRATION NUMBER:  32,430
        REFERENCE/DOCKET NUMBER:  27340-20021.00
    TELECOMMUNICATION INFORMATION:
        TELEPHONE:  650-813-5600
```

```
        TELEFAX:  650-494-0792
        TELEX:  706141
INFORMATION FOR SEQ ID NO:  1003:
    SEQUENCE CHARACTERISTICS:
        LENGTH:  9686 base pairs
        TYPE:  nucleic acid
        STRANDEDNESS:  double
        TOPOLOGY:  circular
        MOLECULE TYPE:  DNA (genomic)
        HYPOTHETICAL:  NO
        ANTI-SENSE:  UNKNOWN
        ORIGINAL SOURCE:
        ORGANISM:  PORYPHYROMONAS GINGIVALIS
    FEATURE:
        NAME/KEY:  misc_feature
        LOCATION:  1..9686
US-09-221-017B-1003
```

```
Alignment Scores:
Pred. No.:                              1.27e-61          Length:             9686
Score:                                  619.00            Matches:            132
Percent Similarity:                     65.72%            Conservative:       54
Best Local Similarity:                  46.64%            Mismatches:         93
Query Match:                            40.78%            Indels:             4
                                        4                 Gaps:               4
DB:

US-09-308-397-2 (1-306) x US-09-221-017B-1003 (1-9686)
```

```
Qy     5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
           |||       :.:||:||||||||||||
Db  4534 GCATTTGTATTCCCCGGTCAGGGAGCACAGTTCGTAGGAATGGGAATGGAAAAGACCTGTACGAA 4475

Qy    25 GlnTyrProIleValLysGluThrIleAspArgAlaSerArgAlaaSerGlnValLeuGlyTyrAspLeu 44
           |||||||  ::: ||||||||||||||||||
Db  4474 CAGAACCCCGAGGCCAAAGCACATATTTGAAAAAAGCCAATGAGATATTAGGCTTTCGCATT 4415

Qy    45 Arg--TyrLeuIleAspThrGluAspLysLeuAsnGlnThrArgTyrThrGlnPro 63
           :::    :::|||:||||
Db  4414 ACGGACATCCTCTTCACGGACACGGCAGAGGAGCTGAAACAAACAAAGTGACCCAGCCG 4355

Qy    64 AlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnLeuLysGlyTyrGlnPro 83
           :::::|||||||  ||| :|::||||
Db  4354 GCGATCTTCCTCACTCGGTGATTCTGGCCCAAGACCGATGGGCGAA--GACTTCCGTCCC 4298

Qy    84 AspMetValAlaGlyTyrLeuSerLeuGlyTyrGluTyrSerAlaLeuAlaSerGlyAlaLeu 103
           |||||||   ||||||:||||||
Db  4297 GATATGGTGGCAGGCCATTCGTCGGAGAGTTTCGGCATTGGTTGCAGCAGGTGCCATG 4238

Qy   104 AspPheGluAspAlaValAlaAlaGluValAlaAlaLysArgAlaTyrMetGluValAlaAla 123
           |||||||  ::: ||||||||
Db  4237 ACCTTCGAAGATGGTCTTCGTCGTATCGAAACGTGCCATGGCCATGAGAAAGCATGC 4178

Qy   124 ProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValAlaValIleGluGlu 143
           |||||||||||||||||:.:||||
Db  4177 GAAGTAGGCCCCTGACTATGCGGCTGTACTGGGGTCGCCTGATGAAAAGTGGAAGAA 4118

Qy   144 AlaCysGlnLysAlaSerGlyValLeuGlyValValThrProAlaAsnTyrAsnThrProAla 163
           ||||||  ::: ||||||||
Db  4117 ATCTGTGCCCAAGTGACCGAT---GAAGTTGTCGTACCGGCCAACTACAACTGCCCGGGA 4061

Qy   164 GlnIleValIleAlaGlyAlaGluValValAlaAlaAspArgAlaValGluLeuGlnGlu 183
           |||||||  :::|||:||
Db  4060 CAGATCGTTATATCCGGTTCCGTAGAAGGAGCTGGAGCTTGCGAACTGCTCAAGGAA 4001

Qy   184 AlaGlyAlaLysArgLeuIleProLeuLysSerGlyProPheHisThrAlaLeuLeu 203
           ||||||  ::::
Db  4000 GCCGGAGCCAAGCGGTGCTTAAGCTCGGTCCGTCGGTTGGAGCATTCCATTCCCCTCGATG 3941

Qy   204 GluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThr 223
           |||||  ::::||||  :::
Db  3940 GAACCGGCTCGTGAAGACTGCCAAAAGCCATCGAGGAGACTACGATCAGTCAGCCGATC 3881

Qy   224 CysProLeuValGlyAsn--ThrGluAlaAlaAlaValMetGlnLysGluAspIleAlaGln 242
```

```
Db   1  ATGAGTAAACAGACAATATTTTTCCGGACAACAAGGTGCCCAAAAAGTTGTATGGCACAA     60
        |||::::||||||                      :::||||        ::||||||
Qy  21  AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu     40
        :::::            :::||||
Db  61  GATTTGTTTAATAACAATGATCAAGCAACTGAAATTTAACTTCGACGCAAGAACGTTA    120
Qy  41  GlyTyrAsp--LeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArg     59
        |||||              ::||||        ::||||
Db 121  GACTTTGATATTTAGAGACAATGTTACTGATGAAGAAGGTAAATTGGGTGAAACTGAA    180
Qy  60  TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGluGlyLys     79
        |||                ||||       :::||
Db 181  AACACGCCACCAGCTTTATTGACGACGATAGTTCGGCATTATTAGCAGCGCTA-----AAA    234
Qy  80  GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla     99
Db 235  ATTTTGAATCCTGATTTACTATGGGCATAGTTTAGGTGAATATTCAAGTTTAGTTGCA    294
Qy 100  SerGlyAlaLeuAspPheGluAspAlaAlaLeuLeuAlaLysArgGlyAlaTyrMet    119
        :::              ||||||   :::||||
Db 295  GCTGACGTATTATCATTTGAAGATGCAGTTAAAAACGTGGTCAATTAATG    354
Qy 120  GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu    139
        ::||||          ||||||   ||||||||
Db 355  GCGCAAGCATTCCTACTGTGTGGAGAAGCATGGCGCAGTATTGGGATTAGATTTTGAT    414
Qy 140  ValIleGluGluAlaCysGlnLys--AlaSerGluLeuGlyValValThrProAlaAsn    158
        :::||||                   ||||||||
Db 415  AAAGTCGATGAAATTGTAAGTCATTATCATCTGATGACAAAATAATTGAACCAGCAAAC    474
Qy 159  TyrAsnThrProAlaGlnIleValIleValIleAlaGlyCysValValAlaAlaArgAlaVal    178
        ||||                    ::||||
Db 475  ATTAATTGCCCAGGTCAAATTGTGTTTCAGGTCACAAAGCTTTAATTGATGAGCTAGTA    534
Qy 179  GluLeuLeuGlnHisAlaGlyAlaGlyLeuIleProLeuLysValSerGlyValProPhe    198
        |||          ||||||          ||||||
Db 535  GAAAAAGGTAAATCATTAGGTGCAAAACGTGTCATGCCTTTAGCAGTATCTGGACCATTC    594
Qy 199  HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer    218
        ||||||          :::             :::
Db 595  CATTCATCCGTAATGAAAGTGATGAAGAGATTTTCAAGTTACATTAAATCAATTTGAA    654
Qy 219  PheSerAspPheThrCysProLeuValGlyAsnThrGluAla--AlaValMetGlnLys    237
        ||||||||          ||||
Db 655  TGGGCTGATGCTAAGTTCCTGTAGTTCAAAATGTAAATGCGCAAGGTGAAACTGACAA    714
Qy 238  GluAspIleAlaGlnLeuLeuThrArgGlnValLysGluThrGlyValArgPheTyrGluSer    257
        :::             :::||||||
Db 715  GAAGTAATTAAATCTAATATGGTCAAGCAATTAATATTCACCAGTACAATTCATTAAACTCA    774
Qy 258  IleGlyValMetGlnLysGlyIleLeuSerAsnPheGluIleGlyProGlyLysVal    277
        |||             :::||||||
Db 775  ACAGAATGGCTAATAGACCAAGGTGTTGATCATTTTATTGAAATTGGTCCTGAAAAGTT    834
Qy 278  LeuSerGlyPheValLysIleAspGlnThrAlaHisLeuAlaHisValGlu    295
        :::             ||||||::::::
Db 835  TTATCTGGCTTAATTAAATAAAAAATAAATAGAGATGTTAAGTTACATCAATTCAA    888
```

RESULT 5
US-09-108-517-1
; Sequence 1, Application US/09108517
; Patent No. 6258934
; GENERAL INFORMATION:
;   APPLICANT: Gentry, Daniel
;   APPLICANT: Lonsdale, John
;   APPLICANT: Pearson, Stewart
;   APPLICANT: Payne, David
;   TITLE OF INVENTION: No. 6258934el FabD
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SmithKline Beecham Corporation
;   STREET: 709 Swedeland Road

```
Db 3880  TGCCCCATCTATCAGAACGTAACGCCGTTCTGCCGTCACCGATCCGGCAGAGATTAAAAAG   3821
         ||||||                     :::||      |||     :::||
Qy 243   LeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGln    262
         :::             ||||||          ::::::
Db 3820  AACCTGATAGCACCGTCACGGCACCGTTCGCTGGACACAGAGTGCTGAAATATGACA     3761
Qy 263   GluAlaGlyIleLeuSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheVal    282
         ::||||::::::||          ::::::
Db 3760  GCCGATGGTGCCGACCACTTTATGGAGTAGGCCCCGGAAATGTATTGCAGGGTTTGGTG   3701
Qy 283   LysLysIle    285
         |||||
Db 3700  AAGAAAATT    3692
```

RESULT 4
US-08-789-609A-1
; Sequence 1, Application US/0789609A
; Patent No. 5827689
; GENERAL INFORMATION:
;   APPLICANT: Gentry, Daniel
;   APPLICANT: Lonsdale, John
;   APPLICANT: Pearson, Stewart
;   APPLICANT: Payne, David
;   TITLE OF INVENTION: No. 5827689el FabD
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SmithKline Beecham Corporation
;   STREET: 709 Swedeland Road
;   CITY: King of Prussia
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19406-0939
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/789,609A
;   FILING DATE: 24-JAN-1997
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/030685
;   FILING DATE: 13-NOV-1996
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Gimmi, Edward R
;   REGISTRATION NUMBER: 38,891
;   REFERENCE/DOCKET NUMBER: GM50004
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 610-270-4478
;   TELEFAX: 610-270-5090
;   TELEX:
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 927 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
US-08-789-609A-1

Alignment Scores:		
Pred. No.:		
Score:	8.15e-55	Length: 927
Percent Similarity:	58.72%	Matches: 547.50
Best Local Similarity:	40.60%	Conservative: 121
Query Match:	36.07%	Mismatches: 54
DB:		Indels: 118
		Gaps: 5

US-09-308-397-2 (1-306) x US-08-789-609A-1 (1-927)

```
Qy   1  MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlnMetGlyArg     20
```

```
;       CITY:        King of Prussia
;       STATE:       PA
;       COUNTRY:     USA
;       ZIP:         19406-0939
; COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette
;       COMPUTER:    IBM Compatible
;       OPERATING SYSTEM: DOS
;       SOFTWARE:    FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/108,517
;       FILING DATE:
;       CLASSIFICATION:
; PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/789,609
;       FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;       NAME:  Gimmi, Edward R
;       REGISTRATION NUMBER: 38,891
;       REFERENCE/DOCKET NUMBER: GM50004
; TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 610-270-4478
;       TELEFAX:   610-270-5090
;       TELEX:
; INFORMATION FOR SEQ ID NO:  1:
; SEQUENCE CHARACTERISTICS:
;       LENGTH:  927 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;       MOLECULE TYPE:  Genomic DNA
US-09-108-517-1

Alignment Scores:
Pred. No.:                              8.15e-55        Length:            927
Score:                                  547.50          Matches:           121
Percent Similarity:                     58.72%          Conservative:       54
Best Local Similarity:                  40.60%          Mismatches:        118
Query Match:                            36.07%          Indels:              5
DB:                                                     Gaps:                4

US-09-308-397-2 (1-306) x US-09-108-517-1 (1-927)

Qy       1  MetThrLysThrAlaPheLeuAlaGlyGlnGlyAlaGlnTyrLeuGlnGlyMetGlyArg  20
            ::: |||
Db       1  ATGAGTAAAACACAGCAATTATTTTTCCGGACAAGGTGCCCAAAAAGTTGGTATGGCACAA  60

Qy      21  AspPheTyrAspGlnTyrProIleValLysGlnValLeuThrIleAspArgAlaSerGlnIleValLeu  40
            |||                                   :::||| ||||||||
Db      61  GATTTGTTTAATAACAATGATCAAGCAACTGAAATTTAACTCAGCAGCAAGAGACGTTA  120

Qy      41  GlyTyrAsp---LeuArgTyrLeuIleAspThrGluGluAspThrLysLeuAsnGlnThrArg  59
            ::: ||| |||
Db     121  GACTTGATATTTTAGAGACAGCAATGTTTACTGATGAAGAAGGTAAATTGGGTGAAACTGAA  180

Qy      60  TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys  79
            |||                                   :::||||||||||||||
Db     181  AACACGCAACCAGCTTATTGACGCATAGTTCGCATTATTAGCAGCGCTA-----AAA  234

Qy      80  GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAla  99
            ||| |||                               ||||||||||||
Db     235  ATTTTGAATCCGATTTTACTATGGGGCATAGTTTAGGTGAATATTCAAGTTTAGTTGCA  294

Qy     100  SerGlyAlaLeuAspPheGluAspAlaIleAlaAlaLeuAlaLysArgGlyAlaTyrMet  119
            ||| |||                               :::||||||||||
Db     295  GCTGACGTATTATCATTTGAAGATGCAGTTAAAATTGTTAGAAAAGGTGGTCAATTAATG  354

Qy     120  GluGluAlaAlaProAlaAspSerGlyMetValAlaValAlaValLeuAsnThrProValGlu  139
            ||| |||                               :::||||||
Db     355  GCGCAAGCAGCATTCCTACTGGTGTAGGAGCGCATGGCTGCCAGTATTGGGATTAGATTTTGAT  414

Qy     140  ValIleGluGluAlaCysGlnLys--AlaSerGluLeuGlyValValValThrProAlaAsn  158
            :::::: |||                            ||||||
Db     130  GACTTTGATTTGCTAGAAACGATGTTTACTGATAATGATGGGAAATTAGGAGAAACAGAA  189
```

```
Db     415  AAAGTCTGATGAAATTGTAAGTCATTATCATCTGATGACAAAATAATTGACCAGCAAAC  474

Qy     159  TyrAsnThrProAlaGlnIleValIleLeuAlaGlyGluValAlaValAlaValAspArgAlaVal  178
            ||| |||
Db     475  ATTAATTGCCCAGGTCAAATTGTTGTTCAGGTACAAAGTTTAATTGATGAGCTAGTA  534

Qy     179  GluLeuLeuGlnGluAlaGlyLysAlaLysProLeuLysValSerGlyProPhe  198
            |||::: |||
Db     535  GAAAAAGTAAATCATTAGGTGCAAAACGTGTCATGCTTTAGCAGTATCTGACCATTC  594

Qy     199  HisThrAlaLeuGluLeuProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer  218
            |||:::::||| |||
Db     595  CATTCATCGCTAATGAAAGTGATTGAAGAAGACGATTTTCAGTTACATTAACAATTTGAA  654

Qy     219  PheSerAspPheThrCysProLeuValGlyAsnThrGluAla--AlaValMetGlnLys  237
            ::: |||
Db     655  TGGCGTGATGCTAAGTTCCTGTAGTTCAAAATGTAAATGCCAAGGTGAAACTGACAAA  714

Qy     238  GluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValAlaArgPheTyrGluSer  257
            ::: ::: |||
Db     715  GAAGTAAATCAATCTAATATGGTCAAGCAATTATATCCACCAGTACAATTCATTAACTCA  774

Qy     258  IleGlyValMetGlnLysAlaGlyIleIleSerAsnPheLeuGluIleGlyLysProGlyLysVal  277
            ::: :::::||| |||
Db     775  ACAGAATGGCTAATAGACCAAGGTGTTGATCATTTTATTGAAATTGGTCCTGGAAAAAGTT  834

Qy     278  LeuSerGlyPheValLysSerIleGluAspGlnThrAlaHisLeuAlaHisLeuValGlu  295
            ::: ::: |||
Db     835  TTATCTGGCTTAATTAAAAAATAAATAGAGATGTTAAGTTAACATCAATTCAA  888
```

```
RESULT 6
US-09-134-001C-1987
; Sequence 1987, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;       APPLICANT: Lynn Doucette-Stamm et al
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;       TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;       FILE REFERENCE: GTC-007
;       CURRENT APPLICATION NUMBER: US/09/134,001C
;       CURRENT FILING DATE:  1998-08-13
;       PRIOR APPLICATION NUMBER: US 60/064,964
;       PRIOR FILING DATE: 1997-11-08
;       PRIOR APPLICATION NUMBER: US 60/055,779
;       PRIOR FILING DATE: 1997-08-14
;       NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1987
;       LENGTH: 936
;       TYPE: DNA
;       ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1987

Alignment Scores:
Pred. No.:                              4.57e-51        Length:            936
Score:                                  515.50          Matches:           116
Percent Similarity:                     58.42%          Conservative:       61
Best Local Similarity:                  38.28%          Mismatches:        111
Query Match:                            33.96%          Indels:             15
DB:                                                     Gaps:                5

US-09-308-397-2 (1-306) x US-09-134-001C-1987 (1-936)

Qy       1  MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlnGlyMetGlyArg  20
            ||| |||                                ||||||||
Db      10  ATGGGTAAAATAGCAGTTGTCTTCCCTGGCCAAGGTGCACAAAAGTAGGGATGGCTAGT  69

Qy      21  AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnIleValLeu  40
            ||| |||                                :::||||||
Db      70  GACTTATATATGAAGAGGAACGCTCGACTGAGTATTAAATTTAGCTCAAGAGACAGTA  129

Qy      41  GlyTyrAspGlyArgTyrLeuIleAspThrGluGluAsp---LysLeuAsnGlnThrArg  59
            ||| |||                                |||
Db     130  GACTTTGATTTGCTAGAAACGATGTTTACTGATAATGATGGGAAATTAGGAGAAACAGAA  189
```

```
;  FEATURE:
;  OTHER INFORMATION: CDC 1551
;  OTHER INFORMATION: "n" bases at various positions throughout the sequence
;  OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:                        1.18e-33           Length:          4403765
Score:                            414.50             Matches:         115
Percent Similarity:               53.47%             Conservative:    39
Best Local Similarity:            39.93%             Mismatches:      117
Query Match:                      27.31%             Indels:          17
DB:                               4                  Gaps:            8

US-09-308-397-2 (1-306) x US-09-103-840A-2 (1-4403765)

Qy       3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe  22
             :::||||||||| ||||||| ||||||||||||||||||||||||||||||||| ::
Db  489058 AAGACGGGCTTTTGTGTTGCCCGGTTCAGGGCTGGGTATGGGTATGGCTAGCGAGCTT  489117

Qy      23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr  42
             ||| || |||||||||||||||||||||||||||||||||||||||||||||||
Db  489118 TATGCTGCCCACCCGGTTTCGCCGAGCCCTCGATGCTGGTCGGAGTTGACCGG  489177

Qy      43 AspLeuArgTyr---LeuIleAspThr--------GluGluAspLysLeuAsnGln  57
             :::|||  ||  ::::
Db  489178 CACCTGCGGTATCGCCTGCGCGATGTGATCTGGGGCCACGACCAAGATCTGTTGAATACC  489237

Qy      58 ThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGln  77
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  489238 ACCGGAATTCGCCCAGCCGCCGTTGTTGCCGGTGGAGTGGCCGTGTATCGGCTCTCATG  489297

Qy      78 GluLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeu  97
             ||  :::  ||
Db  489298 TCGTTGGGGGGTGCGGCCGGTTTGGTGTGCTGGGTCATTCGGTGGGGCGGAGTTGCCGCGGCGG  489357

Qy      98 ValAlaSerGlyAlaLeuAspPheGluAspSerGlyLysMetValAlaValArgGlyAla  117
             ::: |||||||||||||||||||||||||||||||||||||||
Db  489358 CACGTGGTGGGGCCTGTCGTTTGCGCCGGATGCGGCGGGTGGACGG  489417

Qy     118 TyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaLeuAsnThrPro  137
             ||||:::  |||||||| ||||||||||||
Db  489418 TTGATG---CAGGGCGGTGCCGCC---GGCGGCGCCATGTTGCGGTGCAGGCCCGTGAA  489471

Qy     138 ValGluValIleGluGluLeuAlaCysGlnLysAlaSerGluLeuGly----ValValThrPro  156
             |||| ||||:::
Db  489472 GACGAGGTA----------------------GCGGCCGATGCTGGGCACCATGTGAGCATC  489510

Qy     157 AlaAlaAsnTyrAsnThrProAlaGlnIleValIleLeuAlaGlyGluValValAlaValAlaAspArg  176
             ||||  :::|||:::||
Db  489511 GCGGCGGTCAATGGCGGTCCGGTTCGGGTGGGATCCTGGTCGTCGCCACGATGCGGTGAGCGCG  489570

Qy     177 AlaValGluLeuGluGlnGluAlaGlyAlaAlaLysArgLeuIleProLeuLysValSerGly  196
             ::::  ||||||||||
Db  489571 ATCGCTGATCGGCGTGCGCGGCCAGGGC---CGTCGGGTTCCACCGGTTGGCGTCTCGCCAT  489627

Qy     197 ProPheHisThrAlaLeuGluProAlaSerGlnLysLeuAlaLeuThrLeuAlaGln  216
             ||||||||||||||||  :::
Db  489628 GCCTTTCACTCGGCGTTGATGGAGCCGATGATGATCCTGAGTTCACAGGCGTTGGCGGCCGAA  489687

Qy     217 ValSerPheSerAspPheThrCysProLeuValGlyAsn---ThrGluAlaAlaValMet  235
             ||||:::::::
Db  489688 CTGTCTGTGGCTTGCCCACGATCCGGTCGATTCCAATGTGACCGGGCAGTTGGTGGCC  489747

Qy     236 GlnLysGluAspIleAlaGlnIleAlaGlnLeuLeuThrArgGlnLysGluProValArgPheTyr  255
             ||||||||:::  ::::::
Db  489748 GACGACTTCGCCCAGCTGATTACTGGCCGCCGCATATCCGGCGCGTGCGGTTTGGC  489807

Qy     256 GluSerIleGlyIleValMetGlnGluAlaGlyIleSerAspIleSerAsnPheIleGluGlyProGly  275
             ::::||:::::
Db  489808 GACAGTGTTCGTAGTGCCCACTGCGCCGGTGCCAGTCGTTCATCGAAGTGGGCGCCCGGT  489867

Qy     276 LysValLeuSerGlyPheValLys  283
```

Lower alignment section

```
Qy      60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys  79
             ||||| :::
Db     190 AATACACAACCTGCACCATTAACACCATAGTGTTGCGTTACCTAGAATCATTAAAATCAT---  246

Qy      80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuVal  99
             ::: |||
Db     247 ---TTAAAAGTAGAATATACTATCGGGACATAGCTTAGGAGAATATTCAAGCTTAGTTTCT  303

Qy     100 SerGlyAlaLeuAspPheGluAspSerGlyAlaLeuAlaLeuAlaLysArgGlyAlaTyrMet  119
             :::|||:::
Db     304 AGTGATGTATTATCTTTTGAAGATGCGGTTAGAATTGTGCGTAAACCTGGCCAACTTATG  363

Qy     120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGlu  139
             :::|||
Db     364 GCTCAAGCGTTTCCTAACGGTGTTGGGAGTATGGCAGCAGTATTAGGTTTGGATTATGAT  423

Qy     140 ValIleGluGluLeuAlaCysGlnLysAlaSer---GluLeuGlyValValThrProAlaAsn  158
             ::: :::
Db     424 GATGTTGATAAGATATGTCAAACGTTATCTACAAAAGAACAGTTAATTGAACCTGCTAAT  483

Qy     159 TyrAsnThrProAlaGlnIleValIleLeuAlaGlyGluValAlaValAlaAspArgAlaVal  178
             |||||||:::
Db     484 ATTAACTCACCAGGTCAATACTCGTGTGTCTCGAACATAAATCTTTAATTGATGAATTAGTA  543

Qy     179 GluLeuGluGlnGluAlaGlyAlaAlaLysArgLeuIleProLeuLysValSerGlyProPhe  198
             :::||||||:::
Db     544 GAAAAGGCAAAGAACTTGGTGTGCTAAACGTGTTCTTCCATTAGCTGTTTCCGGTCCTTTT  603

Qy     199 HisThrAlaLeuGluProAlaSerGlnLysLeuAlaLeuThrLeuAlaGlnValSer  218
             :::
Db     604 CATTCTTCAATGATGAAAGTTATTGAAGAGAGATTTGCTAATTTCATTAATCAATTGAA  663

Qy     219 PheSerAspPheThrCysProLeuVal----------GlyAsnThrGluAla  232
             ::: |||
Db     664 TGGCATAATGCTAATTATCCAGTTGTTCAGATAATGTTAATGCAAAGGGAGAAACCGATGCT  723

Qy     233 AlaValMetGlnLysGluAspIleAlaGlnIleAlaGlnLeuLeuThrArgGlnLysGluProVal  252
             :::
Db     724 GAAGTAATTAAAACGCAAT----------ATGGTTAAACAATTATATTCACCTGTT  768

Qy     253 ArgPheTyrGluSerIleGlyIleValMetGlnGluAlaGlyIleSerAspIleSerAsnPheIleGluIle  272
             ::: :::
Db     769 CAATTTATTCAATCAACGGAGTGGTTGATTAATCAAGGTGTCGATCACTTTATTGAAATT  828

Qy     273 GlyProGlyLysValLeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAla  292
             :::
Db     829 GGACCGGGAAAAGTATTATCTGGCTTATCATCAAAAAAATAAAATCGAGATGTAAAAATCACT  888

Qy     293 HisValGlu  295
             :::::
Db     889 TCAATTCAA  897
```

RESULT 7
US-09-103-840A-2

```
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT: WHITE, Owen R.
;  APPLICANT: FRASER, Claire M.
;  APPLICANT: VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION: TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE: 1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;  LENGTH: 4403765
;  TYPE: DNA
;  ORGANISM: Mycobacterium tuberculosis
```

```
Db    489868 GGCGGCTTGACGTCGTCGTGATCGAG 489891
                    ||||:::
Db    488129 ATCGGTCGATCGGTCGCCGGCGCCCAGGGA---CGTCGGTCGGTCCACCGGTTGGCGGTTCTCGCAT 488185
                    |||||        :::::::::
RESULT 8
US-09-103-840A-1                                   Qy    197 ProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGlyLeuAlaThrLeuAlaGln 216
; Sequence 1, Application US/09103840A                            :::||||||:||||:||
; Patent No. 6294328                               Db    488186 GCCTTTCACTCGGCGTTGATGGAGCCGATGATCGCTCAGTTCACAGCCGTTGCGCGCCGAA 488245
; GENERAL INFORMATION:                                              ::
; APPLICANT: FLEISCHMAN, Robert D.               Qy    217 ValSerPheSerAspPheThrCysProLeuValGlyAsn---ThrGluAlaAlaValMet 235
; APPLICANT: WHITE, Owen R.                                        :::||:||||
; APPLICANT: FRASER, Claire M.                   Db    488246 CTGTCTGTGGGGTTGCCCAGCGATCCCGGTCATTTCCAATGTGACCGGGCAGTTGGTGCC 488305
; APPLICANT: VENTER, John C.                                      |||                  ||:
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS               Qy    236 GlnLysGluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyr 255
; FILE REFERENCE: 24366-20007.00                 Db    488306 GACGACTTCGCCTCAGCTGATACTGGCCCGGCATATCCGGCGGTGGTGGCGGTTTGGC 488365
; CURRENT APPLICATION NUMBER: US/09/103,840A                      |||        :::::
; CURRENT FILING DATE: 1998-06-24                Qy    256 GluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIleLeuGlyProGly 275
; NUMBER OF SEQ ID NOS: 2                         Db    488366 GACAGTGTTCGTAGTGCCACTGGCGCGGTGCCCAGTCGTTCATCGAAGTCGGGCCCGGT 488425
; SOFTWARE: PatentIn Ver. 2.1                                      ||:||||:||                  |:|||:
SEQ ID NO 1                                       Qy    276 LysValLeuSerGlyPheValLys 283
; LENGTH: 4411529                                                  |||:::               :::
; TYPE: DNA                                       Db    488426 GGCGGCTTGACGTCGTCGTGATCGAG 488449
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv                        RESULT 9
US-09-103-840A-1                                   US-09-428-517-1
                                                   ; Sequence 1, Application US/09428517
Alignment Scores:                                  ; Patent No. 6251636
Pred. No.:              4411529                     ; GENERAL INFORMATION:
Score:                  1.19e-33                    ; APPLICANT: Betlach, Mary C.
Percent Similarity:     414.50                      ; APPLICANT: Shah, Sanjay Krishnakant
Best Local Similarity:  53.47%                      ; APPLICANT: McDaniel, Robert
Query Match:            39.93%                       ; APPLICANT: Tang, Li
DB:                     27.31%                       ; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
                        4                            ; FILE REFERENCE: 30062-20029.00
                                                    ; CURRENT APPLICATION NUMBER: US/09/428,517
US-09-308-397-2 (1-306) x US-09-103-840A-1 (1-4411529)  ; CURRENT FILING DATE: 1999-10-28
                                                    ; EARLIER APPLICATION NUMBER: 60/120,254
Qy      3 LysThrAlaPheLeuPheAlaGlyGlnTyrLeuGlnTyrLeuGlyMetGlyArgAspPhe 22   ; EARLIER FILING DATE: 1999-02-16
                    |||||||||:::::|||:::|||                                  ; EARLIER APPLICATION NUMBER: 60/106,100
Db 487616 AAGACGGGCTTTGTGTTTGCCGGGTCAGGGCTCGCAGTGGTATGGTATGGCTAGCGAGCTT 487675 ; EARLIER FILING DATE: 1998-10-29
                                                    ; NUMBER OF SEQ ID NOS: 12
Qy     23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42  ; SOFTWARE: PatentIn Ver. 2.1
                    ||:|||                     :::|||                        SEQ ID NO 1
Db 487676 TATGCTGCCTACCCGGTTTTCGCCGAGGCCCTCGATGCTGTGGTGGACCAGTTGGACCGG 487735 ; LENGTH: 50937
                                                    ; TYPE: DNA
Qy     43 AspLeuArgTyr---LeuIleAspThr--------GluGluAspLysLeuAsnGln 57     ; ORGANISM: Artificial Sequence
                    ||||||               :::||||                            ; FEATURE:
Db 487736 CACCTGCGGTATCGCGTGCGCGATGTGATCTGGGGCACGACCAAGATCGTTGAAATACC 487795 ; OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
                                                    US-09-428-517-1
Qy     58 ThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGln 77
                    :||                     :::|||:                        Alignment Scores:
Db 487796 ACCGAATTCGCCCAGCCGCGGCTGTTTGCGGTGGAGGTGGCCGCTATCGGCTGCTCATG 487855 Pred. No.:              50937
                                                    Score:                  3.65e-32
Qy     78 GluLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGluTyrSerAlaLeu 97    Percent Similarity:     376.50
                    ::|||                     :::|||:::                    Best Local Similarity:  50.67%
Db 487856 TCGTGGGGGGTGCGGCCGGGGTTGGTGCTGGGTCATTCCGGTGGGCGGAGTTGGCCGGGCG 487915 Query Match:          33.22%
                                                    DB:                     24.80%
Qy     98 ValAlaSerGlyAlaLeuAspPheGluAspPheAlaValAlaLeuValAlaLysArgGlyAla 117                   4
                    :::|||
Db 487916 CACGTCGCGGGGGGCCGGGCTGTGTTCGCCGGATGCGGGTGGCGGGTGGGACGG 487975 US-09-428-517-2 (1-306) x US-09-428-517-1 (1-50937)

Qy    118 TyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro 137  Qy      5 AlaPheLeuPheAlaGlyAlaGlnGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
                    :::|||                     :::|||                        Db   7458 GCCTTTCCTCTTCCCGGGCAGGGCAGCAGCGCCGGCATGGCAGGCGACTCCACCAG 7517
Db 487976 TTGATG---CAGGGCGTTGCCGGCC--GCGGGCGCCATGTTTGCGGCCCGTGAA 488029
                                                    Qy     25 GlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAspLeu 44
Qy    138 ValGluValIleGluGluGlnValAlaCysGlnLysSerAlaSerGluLeuGly---ValValThrPro 156          :::||||
                    ||||||                     :::|||                        Db   7518 GTGTTCCCGGTTCCGGACGCCCTGGACCAGAGGTTGCGCGAACTCGACACCCACCTC 7577
Db 488030 GACGAGGTA---------GCGCCGGATGCTCGGGCACGATGTGAGCATC 488068
                                                    Qy     45 ArgTyrLeuIleAspThrGluGluAspLys------------------- 54
Qy    157 AlaAsnTyrAsnThrProAlaGlnIleValIleValValAlaAlaValAlaAspArg 176               ::|||
                    |||                     :::|||                            Db   7578 GGCCGACTCCTCGCCGCCGAGGCGCCGCCCGAGTGGTTCGCCGAGCGG 7637
Db 488069 GCGGGGGTCAATGGTCCCGGCCTTCGGGCGATCTCGGTGGTCCCCGCCAATGGCGGTGAGCGG 488128
                                                    Qy     55 -------------------------LeuAsnGlnThrArgThrGlnProAlaIleLeuLeuAla 67
Qy    177 AlaValGluLeuGlnGluLeuGlyAlaGlyAlaAlaLysGlyArgLeuIleProLeuLysValSerGly 196
```

```
Percent Similarity:    47.93%              Conservative:   38
Best Local Similarity: 34.83%              Mismatches:    135
Query Match:           24.64%              Indels:         16
DB:                        3               Gaps:            6

US-09-308-397-2 (1-306)  x  US-09-335-409-1 (1-68750)

Qy     3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe  22
           ::::::: :::       ||||||||||||||||||||||||||||||||||||||||
Db 56591 AAGCTCGCCTTCCTGTTCGCCGGCCAGGGCGCAGGGCCATGGGCCGTGGGTTG 56650

Qy    23 TyrAspGlnTyrProIleValLysGluValIleAspArgAlaSerGlnValLeuGlyTyr  42
           :::        :::||||||||||||||||||||||||||||||
Db 56651 TGGGAGGCGGTGGCCGGCGTTCCGGAGACCTTGACCGGTGGTCGTCACCGGTCTTGACCGG 56710

Qy    43 AspLeuArgTyrLeuIle------------------------AspThrGluGluAsp  53
           :::||                                            :::
Db 56711 GAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGGCCAGCCGGGACCGCAGGTCGTCG 56770

Qy    54 LysLeuAsnGlnThrArgTyrThrProAlaIleLeuLeuThrSerValAlaIleTyr  73
           |||||||||||||||||  :::||||||||||| |||||||
Db 56771 TTGCTGGACCAGACGCGCTCACCCGACGGCCGCTCTTGCGCTGGAGTACGGCCGCTGGCC 56830

Qy    74 ArgLeuLeuGluGlnGlyGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu  93
           :::       :::::|||||||||||||||||||||||||||||||
Db 56831 GCGCTCTTCCGGTCGCGGGGCGTGGAGCCGGACCGCGTCGTCGTGCTGGCCATAGCCTCGGCGAG 56890

Qy    94 TyrSerAlaLeuValAlaSerGlyGlyVaLValGluValAlaLeuAspProHisAspAspAla  113
           :::||||||||:::::||||||||||||||||||
Db 56891 CTGGTGGCCGCCTCGCGGGCCGGGGTGTTCTCCTGCGAGGACGGCGTCGGCTTGGTGGTC 56950

Qy   114 LysArgGlyAlaTyrMetGluGluAlaAlaProAlaAspSerGlyLyrValAlaAlaVal  133
           :::         :::|||| |||||||
Db 56951 GCGCGGCGCCGGTGATG---CAGGCGTGCCGGCC---GGCGGCCGGATGGTATCGATC 57004

Qy   134 LeuAsnThrProValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyVal  153
           :::|||||||||| ||
Db 57005 GCCGGCGCCGGAGGCCGACGTG------GCTGCCGCGTGGCGCCACGCAGCGCGTTG 57055

Qy   154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleValAlaGlyGluValValAla  173
           :::     :::|||||||
Db 57056 GTGTCGATCGCCGGCAGTCAATGGGCCGGACGCAGGTGGTCATCGGGGAAAATTC 57115

Qy   174 ValAspArgAlaValAlaGluGluGlnGlnLeuLeuGlnAlaLysArgLeuIleProLeuLys  193
           :::
Db 57116 GTGCAGCAGATCGCGGCGGTTCCGGCGGGGGGCG---CGAACCAAACCGCTGCAT 57172

Qy   194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr  213
           ||||
Db 57173 GTCTCGCACGGCGTTCCACTCGCCGCTCATGGATCCGATCGTGGAGGCGGTTCCGGCGGGTG 57232

Qy   214 LeuAlaGlnIleSerPheSerAspPheCysProLeuValGlyAsnThrGluAlaAla  233
           ||||||| :::::::
Db 57233 ACTGAGTCGGTGACGTACCGGCGCCTTCGCATCGGCTGTGAGCAACCTGAGCCGGAAG 57292

Qy   234 ValMetGlnLysGluAspIleAlaGln--LeuLeuThrArgGlnValLysGluProVal  252
           ||||||||||
Db 57293 CCCTGCACCGATGAGGTGAGCGCGCGGGTTACTCGGGTCGTCACGCGCGAGAGCGCGGTG 57352

Qy   253 ArgPheTyrGluSerIleGlyValValMetGlnGluGluGlyIleSerAsnPheIleGluIle  272
           ||||      :::
Db 57353 CGCTTCGGCGACGAGGTGAAGGGCGTGCAGCGCGGCCGGGTTGCGGGCCTCTTCGTCGAGGTG 57412

Qy   273 GlyProGlyLysValLeuSerGlyPheVal  282
           :::
Db 57413 GGGCCGAAGCCGACCGCTGCTTCCGGCCTTGTG 57442
```

```
Db 7638 GGCACGGGCGCAGCGCCCTGCTTCCCGAGACCCACCACCCAGCGCGCCCCTTCTTCGCC 7697
                                                    |||:::::||   |||||||
Qy   68 ThrSerValAlaIleTyrArgLeuLeuGluGlnGlyTyrGlnProAspMetValAla  87
                                                         :::||
Db 7698 CTGGAAACCGACCCCTTCTTCGCCTTCTGGTCCCAGTGGGCCTGAAACCCGACCACCTCGCA 7757

Qy   88 GlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyGlyValAlaLeuAspPheGluAsp 107
           :::|||||||:::|||||||||||:::||||||||||||||||||||||
Db 7758 GGCCACTCGGTCGGCGATCGCGGCCCACGCAGGCCATCCTCGACCTGTCCGAC 7817

Qy  108 AlaValAlaLeuAlaLysValAlaLysGlyValAlaTyrMetGluGluAlaAlaProAlaAspSer 127
           |||||||||::::::||||||||||||||||||||||
Db 7818 GCGGCCGAACTCGGTGGCCACCCGGCGGCCCGGGCGGTTGATGCGGTTCCCTG------CCCGGCGGC 7871

Qy  128 GlyLysMetValAlaValAlaValLeuAsnThrProValIleGluGluAlaCysGlnLysLys 147
           :::||||||||:::||||||||||||||||||
Db 7872 GGGTCATGTCTCTCGGTCCAGGCACCCGAGTCCGAGTCGCGACACCCTGCTACTCGCGCGT 7931

Qy  148 AlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGlnIleValIle 167
           :::   :::::|||||||||||||||||
Db 7932 GAGGCCCACGTCGGCCTG------GCCGCCGTGACCGGCGGCCGTGGTCGTG 7982

Qy  168 AlaGlyGluValValAlaValAlaAspArgAlaValAlaGluGluGlnGlnLeuLeuGlnAlaLysGlyAlaLys 187
           ||| :::||||||||||||||||||||||||||||||
Db 7983 TCCCGGCGAGCGCGCCAGTCGCGCCGCCATCGAACAGATCCTCGGCCGCAGGGCGCCGCAAA 8042

Qy  188 ArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeuLeuGluProAlaSer 207
           :::||||||||||||||||||||||||||||||
Db 8043 AGCCGGTAC---CTGCCGCTCAGCCACGGCCTTCCACTCCCCGCTCATGGAACCCGTGCTG 8099

Qy  208 GlnLysLeuAlaGluThrLeuAlaGlnIleValSerPheSerAspPheCysProLeuVal 227
           :::        :::|||||||:::::||||||
Db 8100 GAGGAGTTCGGCGAAGAGGCGGTCCGCCTTCGGCAACCACCACCCCCCGTC 8159

Qy  228 GlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAla------GlnLeuLeuThr 245
           :::|||||||||||||||||
Db 8160 TCCAACCTCACCGGCGCCAGTCGACGACCATGGCCGACGTGGCCCGCCTACTGGGTC 8219

Qy  246 ArgGlnValLysGluProValArgPheTyrGluSerIleGlyValValMetGlnGluGluGly 265
           :::|||||||||||||||   :::   ||||||
Db 8220 CGGCACGGTCCGGGAAGCCGGTCCCGCTTCGGCGACGAGGCATCGGGAAACTCGGGC 8279

Qy  266 IleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheValLys 283
           :::||||||||||:::||||||
Db 8280 ACCGGCAGTTCCTGGAGTCGGGACACGGCGTTCCGGCCGACGACCTGCTTCCACCGCGCCATGGCGCGC 8333
```

RESULT 10
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
;   APPLICANT: Schupp, Thomas
;   APPLICANT: Ligon, James
;   APPLICANT: Molnar, Istvan
;   APPLICANT: Zirkle, Ross
;   APPLICANT: Cyr, Devon
;   APPLICANT: Goerlach, Joern
;   TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;   FILE REFERENCE: 4-30582A
;   CURRENT APPLICATION NUMBER: US/09/335,409
;   CURRENT FILING DATE: 1999-06-17
;   NUMBER OF SEQ ID NOS: 30
;   SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

Alignment Scores:
Pred. No.:   1.13e-31   Length: 68750
Score: 374.00   Matches: 101

RESULT 11
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;   APPLICANT: Schupp, Thomas

```
;     APPLICANT: Ligon, James
;     APPLICANT: Molnar, Istvan
;     APPLICANT: Zirkle, Ross
;     APPLICANT: Cyr, Devon
;     APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 1
;     LENGTH: 68750
;     TYPE: DNA
;     ORGANISM: Sorangium cellulosum
US-09-568-102-1

Alignment Scores:
Pred. No.:                        1.13e-31           Length:           68750
Score:                            374.00             Matches:          101
Percent Similarity:               47.93%             Conservative:     38
Best Local Similarity:            34.83%             Mismatches:       135
Query Match:                      24.64%             Indels:           16
DB:                                                  Gaps:             6

US-09-308-397-2 (1-306) x US-09-568-102-1 (1-68750)

Qy        3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
            |||||||||||| |||||||||||||||||||||||||||| |||||||||||||||||||
Db    56591 AAGCTCGCCTTCCTGTTCGCCGGGCAGGGCGCCCAGTGCCCGGGCCGTGGGTG 56650

Qy       23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
            ::|||||||||||||||||| |||||||||| |||
Db    56651 TGGGAGGGGTGGCCGGGCGTTCGCGCAGACCTTCGACCGGTCACGCTCTTCGACCGG 56710

Qy       43 AspLeuArgTyrLeuIle-------------------AspThrGluGluAsp 53
            ::||||||                            :::
Db    56711 GAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGCCCAGCCCGGCCAGCAGGTCGTCG 56770

Qy       54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
            ||||||||||||            |||             |||
Db    56771 TTGCTGGACCAGACGGCGGTTCACCCAGCCGCTCTTTGCCTGGAGTACGCGCTGGCC 56830

Qy       74 ArgLeuLeuGluGluLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu 93
            ||||||                :::||||||||||||::|
Db    56831 GCGCTCTTCGGTGCGTGGGAGCCGGAGCTCGGCTTCGTCGGCGACCGGCTTCGGCGAG 56890

Qy       94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAla 113
            ||:||||||||                    |||||
Db    56891 CTGGTGGGCCGCCTCGGTGGCGTGGGCGGGTGTTCCCCTCCGAGGACGGCGTGGCTGGTC 56950

Qy      114 LysArgGlyAlaTyrMetGluGluValMetGluAlaAlaProAlaAspSerGlyLysMetValAlaVal 133
            ||||||||||              :::|||
Db    56951 GCGCGGCGCGGTGATG---CAGGCGTGCCGGCC---GGCGGCGTGATCGATC 57004

Qy      134 LeuAsnThrProValGluValIleValIleGluLysGlnIleLysAlaSerGluValGlyVal 153
            ||||||                :::
Db    57005 GCCGGCGCCGGAGGCGACGTG-------GCTGCCGGGTGGCGGGCGCCGCACGCAGCGTTG 57055

Qy      154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleIleAlaAlaGlyValValAla 173
            |||
Db    57056 GTGTCGATCGCGCAGTCAATGGGCCGGAGCCGGAGAAATTC 57115

Qy      174 ValAspArgArgAlaValGluLeuGlnGluLeuGluGlyIleGlyAlaAlaLysArgLeuIleProLeuLys 193
            |||   :::
Db    57116 GTGCAGCAGATCGCGGCCGGTTCCACTGCCGCCTCATGGACCGCTGAGGCGGTTCCGGCGGGTG 57172

Qy      194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
            ||| :::
Db    57173 GTCTCGCAGCCGGTTCCACTCGCGCCGCCTCATGGACCGCTGAGGCGGTTCCGGCGGGTG 57232
```

```
Qy      214 LeuAlaLeuGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAla 233
            |||||||||||||||:::::                             |||||||
Db    57233 ACTGAGTCGGTGACGTACCGGCCTTCGATCGCCGTGGTGAGCAACCTGAGCGGGAAG 57292

Qy      234 ValMetGlnLysGluAspPheAlaGln--LeuLeuThrArgGlnValLysGluProVal 252
            |||||||||||||||||||||||      ::|||
Db    57293 CCCTGCACCATGAGGTGACGCGCCGGGTTACTGGGTGCTCACCGCCGAGAGGCGGTG 57352

Qy      253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIle 272
            |||||      ::: ::::                    ||||||:::::
Db    57353 CGCTTCGCCGACGAGTGAAGGCGGTGCACGGCGCCTCCTTCGTCGGAGGTG 57412

Qy      273 GlyProGlyLysValLeuSerGlyPheVal 282
            |||||      |||
Db    57413 GGGGCCGAAGCCGACGCTGCTCGGCCTTGTG 57442

RESULT 12
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;     APPLICANT: Schupp, Thomas
;     APPLICANT: Ligon, James
;     APPLICANT: Molnar, Istvan
;     APPLICANT: Zirkle, Ross
;     APPLICANT: Cyr, Devon
;     APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 1
;     LENGTH: 68750
;     TYPE: DNA
;     ORGANISM: Sorangium cellulosum
US-09-567-969-1

Alignment Scores:
Pred. No.:                        1.13e-31           Length:           68750
Score:                            374.00             Matches:          101
Percent Similarity:               47.93%             Conservative:     38
Best Local Similarity:            34.83%             Mismatches:       135
Query Match:                      24.64%             Indels:           16
DB:                                                  Gaps:             6

US-09-308-397-2 (1-306) x US-09-567-969-1 (1-68750)

Qy        3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
            |||||||||||| |||||||||||||||||||
Db    56591 AAGCTCGCCTTCCTGTTCGCCGGGCAGGGCGCCCAGTGCCCGGGCCGTGGGTTG 56650

Qy       23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
            ::||||||||||||||||||||||||||||||||
Db    56651 TGGGAGGGGTGGCCGGGCGTTCGCGCAGACCTTCGACCGGTCACGCTCTTCGACCGG 56710

Qy       43 AspLeuArgTyrLeuIle-------------------AspThrGluGluAsp 53
            ::||||||                            :::
Db    56711 GAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGCCCAGCCCGGCCAGCAGGTCGTCG 56770

Qy       54 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 73
            ||||||||||||
Db    56771 AAGCTCGCCTTCCTGTTCGCCGGGCAGGGCGCCCAGCAGGGCCGTGGGGCCGTGGGTTG 56830

Qy       23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
            ::||||||||||||||||||||||||||
Db    56651 TGGGAGGGGTGGCCGGGCGTTCGCGCAGACCTTCGACCGGTCACGCTCTTCGACCGG 56710

Qy       43 AspLeuArgTyrLeuIle----------------AspThrGluGluAsp 53
            ::||||||
Db    56711 GAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGCCCAGCCCGGCCAGCAGGTCGTCG 56770

Qy       54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
            ||||||||||||
Db    56771 AAGCTCGCCTTCCTGTTCGCCGGGCAGGGCGCCCAGCAGGGCCGTGGGGCCGTGGGTTG 56830

Qy       74 TyrAspGlnTyrProLeuValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
            ||||||||              ::||||||||||||
Db    56651 TGGGAGGGGTGGCCGGGCGTTCCGGGCCGTTCGACGACCTTCGACCGGTCACGCTCTTCGACCGG 56710

Qy       43 AspLeuArgTyrLeuIle-----------------AspThrGluGluAsp 53
            :::||||||
Db    56711 GAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGCCCAGCCCGGCCAGCAGGTCGTCG 56770

Qy       54 LysLeuAsnGlnThrArgTyrThrGlnProAlaLeuLeuAlaThrSerValAlaIleTyr 73
            ||||||||||||
Db    56771 TTGCTGGACCAGACGGCGGTTCACCCAGCCGCTCTTTGCCTGGAGTACGCGCTGGCC 56830

Qy       74 ArgLeuLeuGluGluLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu 93
            ||||||                :::
Db    56831 GCGCTCTTCGGTGCGTGGGAGCCGGAGCTCGGCTTCGTCGGCGACCGGCTTCGGCGAG 56890

Qy       94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAla 113
            ||:||||||||
Db    57173 GTCTCGCAGCCGGTTCCACTCGCGCCGCCTCATGGACCGCTGAGGCGGTTCCGGCGGGTG 57232
```

```
Qy    3 LysThrAlaPheLeuPheAlaGlyAlaGlnGlyAlaGlnThrTyrLeuGlyMetGlyArgAspPhe  22
Db 56591 AAGCTGCCTTCCTGTTCGCCGGCGCAGGGCGCCAGGTGCCGGGTGCCGGGTG 56650

Qy   23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValAlaGlyTyr  42
Db 56651 TGGGAGGGCGTGGCCGGGCGCCGTTCCGCGAGACCTTCGACCGGTGGCGTCAGCGTCTCGACCGG 56710

Qy   43 AspLeuArgTyrLeuIle-----------------------AspThrGluGluAsp  53
Db 56711 GAGCTCCATCAGCCGCTCGAGGTGATGTGGGCCGGAGGTGCCGGGCCAGCAGCAGTCGTCG 56770

Qy   54 LysLeuAsnGlnThrArgThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr  73
Db 56771 TTGCTGGACCAGACGGCGTTCACCCAGCCGGCGTCTTTGCTGGCGTACGGCTGGCC 56830

Qy   74 ArgLeuLeuGlnGluLysGlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGlu  93
Db 56831 GCGCTCTTCCGGCGTGGGCGTGGCCATAGCCTCGGCGAG 56890

Qy   94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAlaProAlaValAlaLeuValAla 113
Db 56891 CTGTGGCCGCCTGCGTGCGGGGTGTTCTCCCTGGAGACCGCCGTGCGCTTGGTGTC 56950

Qy  114 LysArgGlyAlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaVal 133
Db 56951 GCGCGGCCGGTGATG---CAGGCGCCGCGGCC---GGCGGCGCGATGGTATCGATC 57004

Qy  134 LeuAsnThrProValGluValIleGluGluValIleCysGlnLysAlaSerGluGluValGlyVal 153
Db 57005 GCGCGGCCGGAGGCGACGTG--------GCTGCCGGCGCCCACGCAGCGTTG 57055

Qy  154 ValThrProAlaAsnThrTyrAsnThrProAlaGlnIleValAlaGlyIleGlyValValAla 173
Db 57056 GTGTCGATCGGCAGTCAATGGGCCCGGAGCAGGTGGTGATCGGCGGGCCGAGAAATTC 57115

Qy  174 ValAspArgAlaValGluLeuGlnGluLeuGlnAlaGlyAlaLysGlyLeuIleProLeuLys 193
Db 57116 GTGTCAGCAGATCCGGCGGCGTTCGGGGGCG---CGAACCAAACCGCTGCAT 57172

Qy  194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
Db 57173 GTCTCGCACGCGTTCCACTCGGCGCTCCTCGGATCCGATGCTGAGGCGGTTCCGCGGGTG 57232

Qy  214 LeuAlaGlnValSerPheSerAspPheThrCysProGlyValGlyAsnThrGluAlaThr 233
Db 57233 ACTGAGTCGGTGACGTACCGGCGCCTTGGTGCTGTGGCAACCTGAGCGGGAAG 57292

Qy  234 ValMetGlnLysGluAspPheAlaGln-----LeuLeuThrArgGlnValLysGluProVal 252
Db 57293 CCCTGCACCGATGAGGTGAGCGCGCCGGGTTACGGGCGTCACGGCGAGGAGGCGTG 57352

Qy  253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIle 272
Db 57353 CGCTTCGGCGACGGAGTGAAGGCGGTGAAGGCCGCCGGGCCTTCGTGCGAGGTG 57412

Qy  273 GlyProGlyLysValLeuSerGlyPheVal 282
Db 57413 GGGCCGGAAGCCGAAGCCGACGCGCTGCTCCGGCCTTGTG 57442
```

Alignment Scores:

Pred. No.:	1	Length:	68750
Score:	374.00	Matches:	101
Percent Similarity:	47.93%	Conservative:	38
Best Local Similarity:	34.83%	Mismatches:	135
Query Match:	24.64%	Indels:	16
DB:	4	Gaps:	6

US-09-308-397-2 (1-306) x US-09-568-480-1 (1-68750)

; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; SOFTWARE: PatentIn Ver. 2.0
; NUMBER OF SEQ ID NOS: 30
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-486-1

Alignment Scores:
```
Pred. NO.:                                  Length:        68750
Score:                  1.13e-31            Matches:       101
Percent Similarity:     374.00              Conservative:  38
Best Local Similarity:  47.93%              Mismatches:    135
Query Match:            34.83%              Indels:        16
DB:                     24.64%              Gaps:          6
```

US-09-308-397-2 (1-306) x US-09-568-486-1 (1-68750)

```
Qy       3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db   56591 AAGCTCGCCTTCCTGTTCGCCGGGCAGGGCGCCCAGGGCGCCCAGTACCTGGGCCTGGGTTG 56650

Qy      23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db   56651 TGGGAGGGCGGTGGCCGGCGTTCCGCGAGACCTTCGCGCTCTTCGACCGG 56710

Qy      43 AspLeuArgTyrLeuIle----------------------AspThrGluGluAsp 53
Db   56711 GAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGCCGCAGCGCAGCAGGTCGTCG 56770

Qy      54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
Db   56771 TTGCTGGACCAGACGCGGCGTTCACCCAGCCGGCTGGAGTACGCGCTGGCC 56830

Qy      74 ArgLeuLeuGlnGluLysGlyGlyTyrGlnProAspMetValAlaAlaGlyLeuSerLeuGlyGlu 93
Db   56831 GCGCTCTTCCGGTCTGGGGCGTGGAGCCTGCGCCATAGCCTCGGCGAG 56890

Qy      94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAspAlaLeuAlaLeuValAlaLeuValAla 113
Db   56891 CTGGTGGCCGCCTGCGTGGGTGTGTTCTCCCTGCGCCGTTGGTCGGTGC 56950

Qy     114 LysArgGlyAlaTyrMetGlyAlaProAlaAspSerGlyLysMetValAlaVal 133
Db   56951 GCGGCCGCGCGTTGATG---CAGGCGCTCGGCC---GGCGGCGCGATGGTATCGATC 57004

Qy     134 LeuAsnThrProValGluValIleGluGlnValAlaCysGlnAlaSerGluLeuGlyVal 153
Db   57005 GCGCGGCCGAGGACGCCGTGCGCTTGGTCGGCGTC 57055

Qy     154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleValAlaAlaProAlaSerGlySerGlyValValAla 173
Db   57056 GTGCGATCGGCGAGTCAATGGCGCCGAGCAGGTGGTGTGATCGCGGCGCCGAGAAATTC 57115

Qy     174 ValAspArgAlaValAlaGluGluLeuLeuGlnGluGlnGlyLeuAlaGlyArgLeuIleProLeuLys 193
Db   57116 GTGACCAGACGCGGCGTTCACCCAGCCGGCTGGAGTACGCGCTGCAT 57172

Qy     194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuLeuAlaGlnThr 213
Db   57173 GTCTTCGCACGCGGTTCCACTCGCCCATGGATCCGAGGCGGTTCCGGCGGGTG 57232

Qy     214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAla 233
Db   57233 ACTGAGTCGGTGACGTACCGGCGGCCTTCGATCGCGCGGTGGTGAGCAACCTGAGCGGGAAG 57292

Qy     234 ValMetGlnLysGluProValLysArgGlnValValGlyGluProVal 252
Db   57293 CCCTGCACGCGATGAGGTGAGCCGGCGGCCGCCGGTTACTGGGTGGCCGGAGGGCGGTG 57352
```

; 253 ArgPheTyrGluSerIleGlyGlyValMetGlnGluAlaGlyIleIleSerArgAsnPheGluIle 272
  57353 CGCTTCGGCGACGAGCGGAGTGAAGGCGCTGCACGCGCCGGTGCGGGCCTCTTCGTGCGGAGGTG 57412

Qy   273 GlyProGlyLysValLeuSerGlyPheVal 282
Db   57413 GGGCCGAAGCCCACGCGTGCTCGCGGCCCTGTG 57442

RESULT 15
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;   APPLICANT: Schupp, Thomas
;   APPLICANT: Ligon, James
;   APPLICANT: Molnar, Istvan
;   APPLICANT: Zirkle, Ross
;   APPLICANT: Cyr, Devon
;   APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Alignment Scores:
```
Pred. NO.:                                  Length:        68750
Score:                  1.13e-31            Matches:       101
Percent Similarity:     374.00              Conservative:  38
Best Local Similarity:  47.93%              Mismatches:    135
Query Match:            34.83%              Indels:        16
DB:                     24.64%              Gaps:          6
```

US-09-308-397-2 (1-306) x US-09-568-472-1 (1-68750)

```
Qy       3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe 22
Db   56591 AAGCTCGCCTTCCTGTTCGCCGGGCAGGGCGCCCAGGGCGCCCAGTACCTGGGCCTGGGTTG 56650

Qy      23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr 42
Db   56651 TGGGAGGGCGGTGGCCGGCGTTCCGCGAGACCTTCGCGCTCTTCGACCGG 56710

Qy      43 AspLeuArgTyrLeuIle----------------------AspThrGluGluAsp 53
Db   56711 GAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGCCGCAGCGCAGCAGGTCGTCG 56770

Qy      54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
Db   56771 TTGCTGGACCAGACGCGGCGTTCACCCAGCCGGCTGGAGTACGCGCTGGCC 56830

Qy      74 ArgLeuLeuGlnGluLysGlyGlyTyrGlnProAspMetValAlaAlaGlyLeuSerLeuGlyGlu 93
Db   56831 GCGCTCTTCCGGTCTGGGGCGTGGAGCCTGCGCCATAGCCTCGGCGAG 56890

Qy      94 TyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGluAspAlaLeuAlaLeuValAla 113
Db   56891 CTGGTGGCCGCCTGCGTGGGTGTGTTCTCCCTGCGCCGTTGGTCGGTGC 56950

Qy     114 LysArgGlyAlaTyrMetGlyAlaProAlaAspSerGlyLysMetValAlaVal 133
Db   56951 GCGGCCGCGCGTTGATG---CAGGCGCTCGGCC---GGCGGGCGGCGATGGTATCGATC 57004

Qy     134 LeuAsnThrProValGluValIleGluGlnValAlaCysGlnAlaSerGluLeuGlyVal 153
Db   57005 GCGCGGCCGAGGACGCCGTGCGCTTGGTCGGCGTC
```

Thu Jun 12 07:51:38 2003

us-09-308-397-2.rni

```
Db  57005 GCCGCGCCGGAGGCCGACGTG--------GCTGCCGCGGTGGCCCGACCGCAGCGTTG 57055
                           :::|||               |||         :::

Qy    154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyValValAlaAla 173
              ||:::  | ||             ||:::||||||||||
Db  57056 GTGTCGATCGGGCAGTCAATGGGCGGCAGGTGGTGATCGCCGGCGCCGAGAAATTC 57115

Qy    174 ValAspArgAlaValGluLeuLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLys 193
              |||  :::                       ||||||
Db  57116 GTGCAGCAGATCGGGCGGCGTTCGCGGCGGGGGGGCG---CGAACCAAACCGCTGCAT 57172

Qy    194 ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThr 213
              |||||||        |||:::::||     :::
Db  57173 GTCTCGGCACGCGTTCCACTCGCCGCTCATGATCCGATGCTGGAGGCGTTCCGGCGGGTG 57232

Qy    214 LeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAla 233
              ||::::::          |||  ||||||
Db  57233 ACTGAGTCGGTGACGTACCGGCGGCCTTCGATCGCGCTGGTGAGCAACCTGAGCGGGGAAG 57292

Qy    234 ValMetGlnLysGluAspIleAlaGln--LeuLeuThrArgGlnValLysGluProVal 252
              ||||||                 :::||| |||  |||
Db  57293 CCCTGCACCGATGAGGTGAGCGGCGCGGGGTTACTGGGTGCGTCACGCGCGAGAGGCGGGTG 57352

Qy    253 ArgPheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGluIle 272
              ||||||       :::         |||||:::|||:::
Db  57353 CGCTTCCCGGACGGAGTGAAGGCGCTGCACCCGGCCGTGCGGGGCCTCTTCGTCGTCGAGGTG 57412

Qy    273 GlyProGlyLysValLeuSerGlyPheVal 282
              ||||||           |||
Db  57413 GGGCCGAAGCCGACGCTGCTCGGCCTTGTG 57442
```

Search completed: June 11, 2003, 21:54:52
Job time : 1805 secs

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:    June 11, 2003, 21:03:06 ; Search time 155 Seconds
                                      (without alignments)
                                      2762.958 Million cell updates/sec

Title:           US-09-308-397-2
Perfect score:   1518
Sequence:        1 MTKTAPLFAGQGAQYLGMGR.........QTAHLAHVEDQASLVALLEK 306

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters:     1740770

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO spool/US09308397/runat_06062003_112350_28979/app query.fasta_1.455
-Qf/cgn2_6/ptodata/1/pubppna/PCT_NEW_PUB.seq:* -SUFFIx=rnpb -MINMATCH=0.1
-DB=Published Applications NA -QFMT=fastap -START=1 -END=-1 -MATRIX=blosum62
-LOOPCL=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09308397 @CGN_1_1104_@runat_06062003_112350_28979
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  Published_Applications NA:*
         1: /cgn2_6/ptodata/1/pubppna/US07_PUBCOMB.seq:*
         2: /cgn2_6/ptodata/1/pubppna/PCT_NEW_PUB.seq:*
         3: /cgn2_6/ptodata/1/pubppna/US06_NEW_PUB.seq:*
         4: /cgn2_6/ptodata/1/pubppna/US06_PUBCOMB.seq:*
         5: /cgn2_6/ptodata/1/pubppna/US07_NEW_PUB.seq:*
         6: /cgn2_6/ptodata/1/pubppna/PCTUS_PUBCOMB.seq:*
         7: /cgn2_6/ptodata/1/pubppna/US08_NEW_PUB.seq:*
         8: /cgn2_6/ptodata/1/pubppna/US08_PUBCOMB.seq:*
         9: /cgn2_6/ptodata/1/pubppna/US09_NEW_PUB.seq:*
        10: /cgn2_6/ptodata/1/pubppna/US09_PUBCOMB.seq:*
        11: /cgn2_6/ptodata/1/pubppna/US10_NEW_PUB.seq:*
        12: /cgn2_6/ptodata/1/pubppna/US10_PUBCOMB.seq:*
        13: /cgn2_6/ptodata/1/pubppna/US60_NEW_PUB.seq:*
        14: /cgn2_6/ptodata/1/pubppna/US60_PUBCOMB.seq:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1514	99.7	921	10	US-09-815-242-9482	Sequence 9482, Ap
2	937	61.7	942	10	US-09-815-242-6823	Sequence 6823, Ap
3	920	60.6	3656	10	US-09-070-927A-577	Sequence 577, App
4	643.5	42.4	930	10	US-09-815-242-6028	Sequence 6028, Ap

No.	Length	Score	DB	Pred	ID	Description
5	636	41.9	10	939	US-09-815-242-6878	Sequence 6878, Ap
6	636	41.9	9	1830121	US-09-129-960-1	Sequence 1, Appli
7	627.5	41.3	10	950	US-09-974-300-1143	Sequence 1143, Ap
8	627.5	41.3	10	960	US-09-974-300-1086	Sequence 1086, Ap
9	556.5	36.7	10	906	US-09-815-242-4405	Sequence 4405, Ap
10	556.5	36.7	10	936	US-09-815-242-8032	Sequence 8032, Ap
11	556.5	36.7	10	939	US-09-815-242-7796	Sequence 7796, Ap
12	550.5	36.3	7	6251	US-08-781-986A-25	Sequence 25, Appl
13	421	27.7	10	6021	US-09-070-927A-458	Sequence 458, App
14	416.5	27.4	10	930	US-09-815-242-7193	Sequence 7193, Ap
15	416	27.4	10	592	US-09-974-300-552	Sequence 5552, Ap
16	414.5	27.3	9	4209	US-09-712-363-20	Sequence 20, Appl
17	379	25.0	10	640681	US-09-790-998-1	Sequence 1, Appli
18	376.5	24.8	9	50937	US-09-808-880-1	Sequence 1, Appli
19	374	24.6	9	68750	US-10-014-717-1	Sequence 1, Appli
20	372.5	24.5	9	15872	US-09-860-846-1	Sequence 1, Appli
21	372.5	24.5	9	15872	US-09-988-384B-1	Sequence 1, Appli
22	372.5	24.5	9	15872	US-09-836-821-1	Sequence 1, Appli
23	372.5	24.5	10	15872	US-09-861-289-1	Sequence 1, Appli
24	371	24.4	9	4851	US-09-712-363-116	Sequence 116, Appl
25	348.5	23.0	9	8268	US-10-331-061-83	Sequence 83, Appl
26	348.5	23.0	9	37895	US-10-331-061-1	Sequence 1, Appli
27	348	22.9	9	1380	US-10-074-045-24	Sequence 24, Appl
28	343.5	22.6	9	7959	US-10-331-061-77	Sequence 77, Appl
29	343.5	22.6	9	19227	US-10-331-061-13	Sequence 13, Appl
30	343.5	22.6	10	40138	US-10-331-061-12	Sequence 12, Appl
31	340.5	22.4	9	1278	US-10-124-800-9	Sequence 9, Appli
32	340.5	22.4	9	8730	US-10-124-800-1	Sequence 1, Appli
33	334	22.0	9	1030	US-09-735-056-2	Sequence 2, Appli
34	324	21.3	9	13842	US-09-860-846-30	Sequence 30, Appl
35	324	21.3	9	13842	US-09-988-384B-30	Sequence 30, Appl
36	324	21.3	9	13842	US-09-836-821-30	Sequence 30, Appl
37	324	21.3	10	13842	US-09-861-289-30	Sequence 30, Appl
38	324	21.3	9	36778	US-09-860-846-5	Sequence 5, Appli
39	324	21.3	9	36778	US-09-836-821-5	Sequence 5, Appli
40	324	21.3	10	36778	US-09-861-289-5	Sequence 5, Appli
41	324	21.3	9	37948	US-09-988-384B-5	Sequence 5, Appli
42	270	17.8	9	925	US-09-735-055-1	Sequence 1, Appli
43	254	16.7	9	7692	US-09-836-705-45	Sequence 45, Appl
44	236	15.5	9	34203	US-09-836-705-1	Sequence 1, Appli
45	236	15.5	9	34203	US-09-836-705-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-815-242-9482
; Sequence 9482, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT: Ohlsen, Kari L.
;  APPLICANT: Zyskind, Judith W.
;  APPLICANT: Wall, Daniel
;  APPLICANT: Trawick, John D.
;  APPLICANT: Carr, Grant J.
;  APPLICANT: Yamamoto, Robert T.
;  APPLICANT: Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION: Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
```

```
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   NUMBER OF SEQ ID NOS: 14110
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 9482
;   LENGTH: 921
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(921)
US-09-815-242-9482

Alignment Scores:
Pred. No.:                      1.93e-175
Score:                          1514.00      Length:          921
Percent Similarity:             100.00%      Matches:         305
Best Local Similarity:          99.67%       Conservative:      1
Query Match:                    99.74%       Mismatches:        0
DB:                                          Indels:            0
                                             Gaps:             10

US-09-308-397-2 (1-306) x US-09-815-242-9482 (1-921)

Qy      1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg   20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGACTAAAACAGCCTTTTATTTGCTGGTCAAGGGTGCCCAGTATCTAGGGATGGGACGG   60

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu   40
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GATTTCTATGATCAGTATCCGATTGTCAAAGAAACGATTGATCGAGCGAGTCAGGTGCTA  120

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAspGlnGlnThrArgTyr         60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGTTATGATTTGCGTTATCTCATCGATACGGAAGAGGACAAACTCAATCAGACCCGCTAT  180

Qy     61 ThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuGlnGluLysGlyGly   80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACGCAACCAGCCATTCTAGCGACTTCGGTTGCTATCTACCGTTTATTGCAAGAAAAGGGC  240

Qy     81 TyrGlnProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAlaSer     100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TATCAGCCTGATATGGTTGCTGGTTTGTCTCTTGGAGAATACTCTGCCTTGGTGGCAAGC  300

Qy    101 GlyAlaLeuAspPheGluAspGlyAlaValAlaLeuAlaLysArgGlyAlaTyrMetGlu  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGCGCCCTTGGATTTTGAAGATGCGGTTGCCTGGTAGCGTGGGAGCCTATATGGAA      360

Qy    121 GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal  140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAAGCGGCTCCTGCTGACTCGGCAAGATGGTAGCAGTTCCAATACGCCAGTAGAGGTC    420

Qy    141 IleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsn  160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATTGAAGAAGCCTGTCAAAAAGCTTCTGAACTTGGAGTGGTTACTCCAGCCAACTATAAC  480

Qy    161 ThrProAlaGlnIleValIleAlaGlyGluValAlaAlaAlaValAspArgAlaValGluLeu  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACACCTGCACAAATCGTCATTGCTGGAGAAGTGGTTGCAGTGACGAGGGGTTGAACTT    540

Qy    181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThr  200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTGCAAGAAGCAGGTGCCAAACGCTTGATTCCTCTTAAGGTGTCAGGTCCCTTTCACACC  600

Qy    201 AlaLeuGluGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnInValSerPheSer  220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCTCTCCTTGAACCTGCTAGCCAGAAACTAGCTGAAACTCTAGCTCAGGTAAGTTTTTCA  660

Qy    221 AspPheThrCysProLeuValGlyAsnThrGlnAlaValMetGlnLysGluAspIle      240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GATTTTACTTGTCCCCTAGTCGGCAATACAGAAGCTGTCGTGATGCAAAAAGAGAGACATT  720
```

```
Qy    241 AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal  260
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCTCAGCTCTTGACGCGTCAGGTCAAGGAACCCGTTCGTTTCTATGAAAGTATTGGGGTC  780

Qy    261 MetGlnGluAlaGlyIleIleSerAsnPheGluGluIleGlyProGlyLysValLeuSerGly  280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ATGCAAGAGGCAGGCATAAGCAACTTTATCGAGATTGGACCGGGGAAGTTTGTCAGT    840

Qy    281 PheValLysLysIleAspGlnThrAlaHisIleLeuAlaHisValGlyAspGlnAlaSerLeu  300
            :|||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db    841 TTTGTTAAAAAAATTGATCGAACTGCTCACTTAGCTCACGTCATGTGGAAGATCAAGCGAGTTTA  900

Qy    301 ValAlaLeuLeuGluLys  306
            ||||||||||||||||||
Db    901 GTAGCACTTTTAGAAAAA  918
```

```
RESULT 2
US-09-815-242-6823
;   Sequence 6823, Application US/09815242
;   Patent No. US2002061569A1
;   GENERAL INFORMATION:
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari L.
;   APPLICANT: Zyskind, Judith W.
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John D.
;   APPLICANT: Carr, Grant J.
;   APPLICANT: Yamamoto, Robert T.
;   APPLICANT: Xu, H. Howard
;   TITLE OF INVENTION: Identification of Essential Genes in
;   TITLE OF INVENTION: Prokaryotes
;   FILE REFERENCE: ELITRA.011A
;   CURRENT APPLICATION NUMBER: US/09/815,242
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: 60/191,078
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: 60/206,848
;   PRIOR FILING DATE: 2000-05-23
;   PRIOR APPLICATION NUMBER: 60/207,727
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: 60/242,578
;   PRIOR FILING DATE: 2000-10-23
;   PRIOR APPLICATION NUMBER: 60/253,625
;   PRIOR FILING DATE: 2000-11-27
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   NUMBER OF SEQ ID NOS: 14110
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 6823
;   LENGTH: 942
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(942)
US-09-815-242-6823

Alignment Scores:
Pred. No.:                      5.94e-105
Score:                          937.00       Length:          942
Percent Similarity:             75.17%       Matches:         187
Best Local Similarity:          61.92%       Conservative:     40
Query Match:                    61.73%       Mismatches:       75
DB:                                          Indels:            0
                                             Gaps:              0

US-09-308-397-2 (1-306) x US-09-815-242-6823 (1-942)

Qy      3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe   22
            |||:|||:||||||:||||||:||||||||||||||||||||||:|:|||:
Db     16 AAACAGGCGATTTTATTTAGTGGACAAGGAGCCCAGTATCAGGGATGGGTGAAGAATTA   75
```

```
Qy    23  TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr  42
Db    76  TATCACCAAGAACGCGATTGTTCGGGAAACTTTCGATGAGCAAGTCATATCTTAGGTTAT  135

Qy    43  AspLeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGln  62
Db   136  GAGATGGCAGAACTTTGTTTACTGAAAATGAACGTTAAATGAACAGAATATACGCAA  195

Qy    63  ProAlaIleLeuAlaThrSerValAlaIleAlaThrGlyLeuGlnLysGlyTyrGln  82
Db   196  CCTGCTATTTTAACAGTCAGTGTCGCATTTACCGTCTTTTGCAACAAAAGGACTAACG  255

Qy    83  ProAspMetValAlaGlyLeuSerLeuGlyTyrSerAlaLeuValAlaSerGlyAla  102
Db   256  CCTGATGTCGTAGCGGGGTTTAAGCTAGGGGAATATAGCGCTTTGGTTGCCAGCGGGGCT  315

Qy   103  LeuAspPheGluAlaValAlaAlaLeuValAlaLysArgGlyAlaTyrMetGluGluAla  122
Db   316  TTGGCCTTTCAGAAGCAGTGGCCTTGGTCCAAAAGCGCGGTCAGTACATGACAGAAGCA  375

Qy   123  AlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVallleGlu  142
Db   376  GCACCACAAGGAACTGGCAAAATGGTTGCTGTCGTAAATGCTGAGCGTGAAGTAATTGAG  435

Qy   143  GluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrPro  162
Db   436  AAAGCCTGCCAAGAAGCCAGTGCTTCGGAAATTGTGGCTCCAGCAAATTATAATACACCA  495

Qy   163  AlaGlnIleValIleAlaGlyGlyValValAlaValAspArgArgValAlaGluLeuLeuGln  182
Db   496  CAACAAATCGTGATTGGTGGTGAGGTTGCTGCTGTTGATCAAGGATGACACTTCTCAAA  555

Qy   183  GluAlaGlyAlaAlaLysGlyAlaLeuIleProLeuLysValSerGlyProPheHisThrAlaLeu  202
Db   556  GAAGCTGGTGTGAAGCGAATGATTCCGTTAAATGTGAGTGGCCCTTTCCATACGGCGCTG  615

Qy   203  LeuGluProAlaSerGlnValAlaSerGluThrGlnLeuAlaGlnValSerPheSerAspPhe  222
Db   616  TTACAACCAGCATCAAAAAAATTGGCTCAGGATTTAGCAAAAATTGAACTTTCAAAACGATG  675

Qy   223  ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAspIleAlaGln  242
Db   676  GCTATGAACGTGGGAAACGATGATTGAAGTTGGTCCAGGGAAAACATTAACTGGTTTTGTT  735

Qy   243  LeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGln  262
Db   736  TTATTGGAAAAGCAAGTCATGTCGCGGTACGTTTTGAAGACAGTATCGAAACGATGAAG  795

Qy   263  GluAlaGlyValIleSerAsnPheIleGlulleGlyLeuGlyProGlyLysValLeuSerGlyPheVal  282
Db   796  GCTATGAACGTGGGAAACGATGATTGAAGTTGGTCCAGGGAAAACATTAACTGGTTTTGTT  855

Qy   283  LysLysIleAspGlnThrAlaHisLeuAlaHisValGluValGlnAlaSerLeuValAla  302
Db   856  AAAAAAATTGACAAACAATTGAAATGCACCGTGTGGAAGATGTTGCAACATTAACAGAA  915

Qy   303  LeuLeu  304
Db   916  ACGTTA  921
```

RESULT 3
US-09-070-927A-577
; Sequence 577, Application US/09070927A
; Patent No. US2002012011GA1
; GENERAL INFORMATION:
    APPLICANT: Charles A. Kunsch
             Patrick J. Dillon
             Steven Barash
    TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
    NUMBER OF SEQUENCES: 982
    CORRESPONDENCE ADDRESS:
       ADDRESSEE: Human Genome Sciences, Inc.
       STREET: 9410 Key West Avenue

```
;: LOCATION: (1)...(930)
US-09-815-242-6028

Alignment Scores:
Pred. No.:
Score:                          4.1e-69
Percent Similarity:             643.50
Best Local Similarity:          64.84%
Query Match:                    45.48%
DB:                             42.39%
                                10

Length:                         930
Matches:                        141
Conservative:                   60
Mismatches:                     102
Indels:                         7
Gaps:                           4
```

```
US-09-308-397-2 (1-306) x US-09-815-242-6028 (1-930)

Qy    1  MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
            :::|||:|  ::  :||||  :|  ||||  |  ::: || |    :|
Db    1  ATGACGGCAATTTGCATTTGTGTTCCCTGGACAGGGTTCTCAAACCGTTGAATGCTGGCT  60

Qy   21  AspPheTyrAspGlnTyrProIleValSerGluThrIleAspArgAlaSerGlnValLeu  40
            : |||  ||||||  ||||  | ::|||| ||||::|| || ::|||
Db   61  GATATGGCGGCAGGCTATCCAATTGTGAAGAAACGTTTGCTGGAGCTTCGGCGCGCTG  120

Qy   41  GlyTyrAspLeuArgTyrLeuIleAspThrGluGlu------------AspLysLeuAsnGln  57
            ||||||  |||  ||||
Db  121  GGCTACGACCTGTGGGCGTG------ACCCAGCAGGGGCCAGCTGAAGAACTGAATAAA  174

Qy   58  ThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGln  77
            :::|| || ||:|| |:|||  | ::|  :|||  |  ::|
Db  175  ACCTGGCAAACTCAGCCTGCCATCCTGGCTGTGCGTGATCGTCGTATGGGCAG  234

Qy   78  GluLysGlyTyrGln---ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAla  96
            :: ||||    :    ||||  ||:|| :|  ||||||  :||
Db  235  CAGCAGGGCCGGTAAAGCCAATGATGGCCGTCACAGCCTGGGGAATACTCCGCG  294

Qy   97  LeuValAlaSerGlyValAlaLeuAspPheGluAspGlyAlaAspAlaAlaLysArgGly  116
            :::|| |||  ||  | |   |||| :|  |||   |   || :
Db  295  CTGGTTTGCGCTGGTGTGATTGATTTCGCTGATGCGTCTGGTTGGATGCGCGGGC  354

Qy  117  AlaTyrMetGluGluGluArgProAlaAlaAspSerGlyLysMetValAlaValLeuAsnThr  136
            ::|||  :|  |||   |   |   ::|  |  :| :|||
Db  355  AAGTTCATGCAAGAAGCCGTACCGAAGGCACGGGCGCTATGGGCGCAATCATCGGTCTG  414

Qy  137  ProValGluValIleGluGluGlnAlaCysGlnLysAlaSerGluLeuGlyValValThrPro  156
            ||:   : || :||||| | | ||   : | :|||
Db  415  GATGATGGCGTCTATTGCGAAGACGTGTGGAAGAACTGAGAAGGTCAGGTCGTTTCTCCG  474

Qy  157  AlaAsnTyrAsnThrProAlaGlnIleValIleLeuAlaGlyValGluValAlaValAspArg  176
            ::|||   :    :|  ||||||:|  | |  :|||:
Db  475  GTAAACTTTAACTCTCCGGACAGCAGTGGTTATTGCCGGTCATAAAGAAGCGGTTGAGCGT  534

Qy  177  AlaValGluLeuGlnGluLeuGlnAspGlyAlaGlyArgLeuIleProLeuLysValSerGly  196
            ::|||  || | || | :||||  :| ::|||| :|
Db  535  GCTGGCGCTGCCTGTAAAGCGGGCGGGCGCAAAACCGCGCCGTCCGTTACCAGTGAGCGTA  594

Qy  197  ProPheHisThrArgLeuGluLeuProAlaSerGlnLysLeuThrLeuLeuAlaGln  216
            :: |||  ::  ::||||  :|  || ::|
Db  595  CCGTCTCACTGTGCCGTGATGAAACCAGCAGCCAGCAAACTGGCAGTAGAATTAGCGAAA  654

Qy  217  ValSerPheSerAspPheThrCysProLeuValGlyAsnThrGluAlaAlaValMetGln  236
            :||  | |||  |||  : :| ::|  ||||
Db  655  ATCACCTTTAACGCACCAACAGTTCCTGTTGTGAATAACGTTGATGTGAAATGGCGAAACC  714

Qy  237  GluGluAsp---IleAlaGlnLeuLeuThrArgGlnValValGluProValArgPheTyr  255
            :::|||  ::       :||| :   ::
Db  715  AATGGTGATGCCATCCGTGACGCACTGGTACGTCAGTTGTATAACCCGGTTCAGTGACG  774

Qy  256  GluSerIleGlyValMetGlnGlnValGlyIleLeuSerAsnPheIleGluIleGlyProGly  275
            :: :|||   :|     :::   | ::
Db  775  AAGTCTGTTGATGAGTACATGGCAGGCAAGGCTAGAACATCTCATGAAGTGCGGCCCGGGC  834

Qy  276  LysValLeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuValGlu  295
            :: ||||   |||   :::|:
Db  835  AAAGTGCTTACTGGCTGACGAAACGCATTGTCGACACCCTGACGCCGCTGAAC  894
```

```
Qy  143  GluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAlaAsnThrPro  162
            :::|||  :  ||| ||||||
Db 3173  AAAGCCTGCCAAGGCCAGTGCTTTCGGAATTGTGGCTCCAGCAATTATAATACACCA  3232

Qy  163  AlaGlnIleValIleLeuAlaGlyValAlaValAlaValAspArgAlaValGluLeuGln  182
            |||  :| ||||| :|  ||| ||||  :::
Db 3233  CAACAAATCGTGATCCTGGCTGGTGTGGAAGTGGCTGTGATCAAGCGATGACACTTCTCAA  3292

Qy  183  GluAlaGlyAlaGlyArgLeuIleProLeuLysValSerGlyProPheHisThrAlaLeu  202
            :::|| ||||||    :  || :|  ||:::
Db 3293  GAAGCTGGTGCTGGCGAATGATTCCGTTAAATGTGAGTGGCCCTTTCCATACGGCGCTG  3352

Qy  203  LeuGluProAlaSerGlnLysLeuThrLeuLeuAlaGlnValSerPheSerAspPhe  222
            :::||   :  |||  :||:
Db 3353  TTACAACCAGCATCAAAAAAATTGGCTCAGGATTTAGCAAAATTGAACTTTCAAACGATG  3412

Qy  223  ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnGluGluAspIleAlaGln  242
            :| ::|  |||  :  :::  ::
Db 3413  CAAATTCCTGTCATTAGTACAGACTGCGAAATTATGCCCAAGAGGCAATTCAAGCGACG  3472

Qy  243  LeuLeuThrArgGlnValValGluProValArgPheTyrGluSerIleGlyValMetGln  262
            :||||    :: | :
Db 3473  TTATTGGAAAACGCAAGTCATGTCTCGGTAGGTTTGAAGACAGTATCGAAACGATGAAG  3532

Qy  263  GlyAlaGlyIleLeuSerAsnPheIleGluIle-GlyProGlyLysValLeuSerGlyPheVal  282
            :: |||   |  ||: ::|
Db 3533  GCTATGAACGTAGGAACGATGATTGAAGTTNGGTCCAGGGRAAACATTAACYGGTTTTGT  3592

Qy  282  llysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu  300
            :: :::  :
Db 3593  TAAAAAAATTGACCAAACAATTGAAATGCACCGTGTGGAAGATGTGTTGCCACATTA  3647
```

```
RESULT 4
US-09-815-242-6028
; Sequence 6028, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6028
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
```

```
Qy    296  AspGlnAlaSerLeuValAlaLeuGlu 305
           :::      ||||||| |||||
Db    895  GAACCTTCAGCGATGGCAGCGCGGCGCTCGAG 924

RESULT 5
US-09-815-242-6878
; Sequence 6878, Application US/09815242
; Patent No. US2002006156941
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA_011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6878
;   LENGTH: 939
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(939)

US-09-815-242-6878

Alignment Scores:
Pred. No.:                      3.43e-68
Score:                          636.00
Percent Similarity:             63.43%
Best Local Similarity:          45.95%
Query Match:                    41.90%
DB:                             10
```

```
US-09-308-397-2 (1-306) x US-09-815-242-6878 (1-939)

Qy      1  MetThrLysThrAlaPheLeuPheAlaGlyAlaGlnTyrLeuGlyAlaGlnTyrLeuGlyMetGlyArg 20
           |||                      ||  |  :|:|||| :|:||||
Db      1  ATGAAAAAATTCGCAATGGTCTTCCCAGGTCGTCAAGGCTCCCAAACTGTCGGTATGCTTGCT 60

Qy     21  AspPheTyrAspGlnTyrProIleValLysGluThrIleLysAspAspArgAlaSerGlnIleValLeu 40
           |||              |||||||        ||::||
Db     61  GATCTTGCAACTGAATATCCAATCGTTATTGAAACATTTAAACAAGCATCTGATGCGCTT 120

Qy     41  GlyTyrAspLeuArgTyrLeuIleAspThr---GluGluLysLeuAsnGlnThrArg 59
           |:|              |:     :|:|||||||||||
Db    121  GGTTATGATTATGGTATCTTGTTCAACAAGGTCCAGCTGAAGAACTTAATAAAACTGG 180

Qy     60  TyrThrGlnProAlaLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys 79
           |:|:::|||:|::  |:|||||| |:|:||||| ||||||
Db    181  CAAACTCAGCCCCCACTTTTAGCTGCTTCAGTCGCTATTTATCGCTATTTATGGCGTATGAAAGAAAAA 240
```

```
Qy     80  GlyTyrGln------ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeu 97
           :||||||          ||||||||||||| ||||| ||| |||||||||
Db    241  TTTCCTCAATTAAAACCAGAAGTGATGGCAGGTCAGGTCAGTGAGATTCTGCGTTA 300

Qy     98  ValAlaSerGlyAlaLeuAspPheGluAspArgAlaLeuAlaLeuAlaLysArgGlyAla 117
           |||      :||||| :||||:::|||||  ||||||||||||||
Db    301  GTTTGTGCTGGCGTGTTGGATTTCCAAGATCGCGATTAAAATTAGTGAAATTGCCGCGAAAA 360

Qy    118  TyrMetGluGluAlaIleAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro 137
           |:|      :||||   |:|||::||||:|||||||||||
Db    361  TTAATGCAACAGCTGTGCCTGAAGGCACTGGCAATCATTGGTTAGAT 420

Qy    138  ValGluValIleGluGluAlaCysGlnLysAlaSerGluLeuGlyValValThrProAla 157
           |||:  ||||||:|||||||  |:||||||||||||||||
Db    421  AATGAAGCAATTATTAATGCTTGCAACAGCAGGAAGGCGAAGTCGTATCTGCGGTG 480

Qy    158  AsnTyrAsnThrProAlaGlnIleValIleGluAlaGlyValValAlaValAlaAspArgAla 177
           |:::|:::|||:|||||  ||| :|:::|||||||||||
Db    481  AACTTTAACTCACCGGGTCAAGTAGTTATTGCGGTGGCGAAAGCGCAGTGAGCCTGCG 540

Qy    178  ValGluLeuLeuGlnGluAlaGlyAlaArgLeuIleProLeuLysProValSerGlyPro 197
           |||                 :::||||||||| ||||||||
Db    541  GCTGCATTATGTAAAGAAGCAGGGGCGCAAACTGCATTGCCGTTAGCTGTGAGCGTACCT 600

Qy    198  PheHisThrAlaLeuGluGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnVal 217
           |||              :||||:   ||||||||||||:  :::
Db    601  TCTCACTGTGCATTAATGAAACCTGAGCCGGCCAATTAGCGGTAACACTTGAGAATATT 660

Qy    218  SerPheSerAspPheThrCysProLeuValGlyAsnThrGluAla---AlaValMetGln 236
           |||                |:|     ||||||||||||
Db    661  CAAATTAATACACCAACAATATCGGTATTAAATAACGTTGATGTGAAAGCTGAAACTGAA 720

Qy    237  LysGluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGlu 256
           |||              |::     :|:||||||||||
Db    721  GGCACCGAAATTCGTACCGGCACTTGTGCGTCAGTTATATAGTCCAGTTCGTTGGACTGAA 780

Qy    257  SerIleGlyValMetGlnGlnValAlaGlyIleSerAsnPheIleGluIleGlyProGlyLys 276
           |||              :::   ||||||||||||| :::|:
Db    781  ACAGTTGAAAAAATGGCCAAGATGGCGTTCTAGTTGTTGCTGTTCTCTGAAGTGGGGCCAGGTAAA 840

Qy    277  ValLeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAsp 296
           |||              |||  :::||  |||||||||||
Db    841  GTATTAAATGGTTTAACCAAACGCATTGTGGGTGATTTACAAGCAATATCTGTAAATGAT 900

Qy    297  GlnAlaSerLeuValAlaLeuLeuGlu 305
           ||||||          ::|: |||
Db    901  GTTGCATCATTCAATGCGGTAGAAGAA 927

RESULT 6
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fra
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 1830121
;   TYPE: DNA
;   ORGANISM: Haemophilus influenzae
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
Length:            939
Matches:           142
Conservative:       54
Mismatches:        109
Indels:              4
Gaps:                3
```

LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (123336)..(123336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (142750)..(142750)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: -(145058)..(145058)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145171)..(145171)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (145942)..(145942)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (147197)..(147197)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (150841)..(150841)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152500)..(152500)
;   OTHER INFORMATION: n equals a, t, g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.:                        2.37e-63          Length:              1830121
Score:                            636.00            Matches:             142
Percent Similarity:               63.43%            Conservative:        54
Best Local Similarity:            45.95%            Mismatches:          109
Query Match:                      41.90%            Indels:              4
DB:                                                 Gaps:                3

US-09-308-397-2 (1-306) x US-10-329-960-1 (1-1830121)

Qy     1  MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
          ::: ::|  |  || |||   |||     ||::
Db 172507  ATGAAAAAAATTCGCAATGG·CTTCCCAGGTCAAGGCTCCCAAACTGTCGGTATGCTTGCT 172448

Qy    21  AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
          |||||||||||  |||  ||||   || |||     ::
Db 172447  GATCTTGCAACTGAATATCCAATCGTTATTGAAACATTAAACAAGCATCTGATGCGCTT 172388

Qy    41  GlyTyrAspLeuArgTyrLeuIleAspThr---GluGluAspLysLeuAsnGlnThrArg  59
          ::::|  |||||||   ||||  :||      ::::::::
Db 172387  GGTTATGATTGATTTATGTGTCAACAGGTCCAGCTGAAGAACTTAATAAACTTGG 172328

Qy    60  TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys  79
          |||   |||  :::||||||
Db 172327  CAAACTCAGCCCGCACTTTTAGCTCGTTCAGTCGCTATTTATCGCGTATGGAAAGAAAAA 172268

Qy    80  GlyTyrGln------ProAspMetValAlaGlyTyrLeuSerLeuGlyGluTyrSerAlaLeu  97
          |||  ||           :::::::|| ::
Db 172267  TTTCCTCAATTAAAACCAGAAGTGATGGCAGGTCATAGTGAGTGATTCTGCGTTA 172208

Qy    98  ValAlaSerGlyAlaLeuAspPheGluAspAlaVaIAlaLeuValAlaLysArgGlyAla  117
          |||||||||  ::  |||:::
Db 172207  GTTTTGTCTGCGTGTTGGATTTCCAAGATGCAATAAATTAGTGAATTGCGCGGAAAA 172148

Qy   118  TyrMetGluGluGluAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro  137
          |||:::::
Db 172147  TTAATGCAACAAGCTGTGCCTGAAGGCACTGGCCAATGTATGCAATCATTGGTTTAGAT 172088

Qy   138  ValGluValIleGluGluValAlaCysGlnLysLysAlaSerGluLeuGlyLeuGlyValValThrProAla  157
          ||||||  ||||||||
Db 172087  AATGAAGCAATTATTAATGCTTGCAAACAAGCAGAGGAGGCGAAGTCGTATCTGCGGTG 172028
```

```
Qy   158  AsnTyrAsnThrProAlaGlnIleValIleAlaGlyGluValValAlaValAspArgAla  177
          ||:::|  || |||||||
Db 172027  AACTTTAACTCACCGGGTCAAGTAGTTATTGCGGGTGCGAAAGTTGCTGAGCGGTGCG 171968

Qy   178  ValGluLeuGlnGlyAlaGlyLysArgLeuIleProLeuLysValSerGlyPro  197
          |||  ::  || |  ||||||
Db 171967  GCTGCATATGTAAAGAACGAGCAGGGCGAAACGTCATTGCCGTTAGCTGTGACGTACCT 171908

Qy   198  PheHisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnVal  217
          |||||::::|
Db 171907  TCTCACTGTGCATTAATGAAACCTGCAAGCCAGCAATTAGCCGTAACACTTGAGAATATT 171848

Qy   218  SerPheSerAspPheThrCysProLeuValGlyAsnThrGluAla--AlaValMetGln  236
          ||||  :::  :::
Db 171847  CAAATTAATACACCACCAATCATATCGTTATTAAATAACGTTGATGTGAAAGCTGAAACTGAA 171788

Qy   237  LysGluAspIleAlaGlnLeuLeuThrArgGlnValValTyrGluProValArgPheTyrGlu  256
          ||  :::  |  |||
Db 171787  GGCACCGAAATTCGTACCGCACTTGTGCGTCAGTTATATAGTCCAGTTCGTTGGACTGAA 171728

Qy   257  SerIleGlyValMetGlnGluAlaGlyIleIleSerAsnPheIleGluIleGlyProGlyLys  276
          :::::|||  :::
Db 171727  ACAGTTGAAAAAATGGGCCAAGATGGCGTTCCAGTCGTTGCTTGAAGTGGGGCCCAGGTAAA 171668

Qy   277  ValLeuSerGlyPheValLysIleGluAspGlnThrArgAlaHisLeuAlaHisValGluAsp  296
          :::||  ||
Db 171667  GTATTAAATGGTTTAACCAAACGCATTGTGGGTGATTTACAACAAGCAATATCTGTAAATGAT 171608

Qy   297  GlnAlaSerLeuValAlaLeuLeuGlu  305
          :::  :|  :::
Db 171607  GTTGCATCATTCAATGCGGTAGAAGAA 171581
```

```
RESULT 7
US-09-974-300-1143
; Sequence 1143, Application US/0974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;   APPLICANT: Berka, Randy M.
;   APPLICANT: Clausen, Ib Groth
;   TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;   TITLE OF INVENTION: Expression
;   FILE REFERENCE: 10085.500-US
;   CURRENT APPLICATION NUMBER: US/09/974,300
;   CURRENT FILING DATE: 2001-10-05
;   PRIOR APPLICATION NUMBER: 09/680,598
;   PRIOR FILING DATE: 2000-10-06
;   PRIOR APPLICATION NUMBER: 60/279,526
;   PRIOR FILING DATE: 2001-03-27
;   NUMBER OF SEQ ID NOS: 8481
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 1143
;   LENGTH: 950
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-1143

Alignment Scores:
Pred. No.:                        3.81e-67          Length:              950
Score:                            627.50            Matches:             139
Percent Similarity:               61.87%            Conservative:        46
Best Local Similarity:            46.49%            Mismatches:          111
Query Match:                      41.34%            Indels:              3
DB:                                                 Gaps:                3

US-09-974-300-1143 (1-306) x US-09-974-300-1143 (1-950)

Qy     1  MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
          |||  |||     ||||  ||  |||
Db     1  ATGGGCAAGATTGCTTTTCATTCCCGGGCCAAGGTTCGCAGCATATCGGCCATGGCGACAC  60

Qy    21  AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
          :::::::
```

```
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1086

Alignment Scores:
Pred. No.:              3.8e-67        Length:           960
Score:                  627.50         Matches:          139
Percent Similarity:     61.87%         Conservative:      46
Best Local Similarity:  46.49%         Mismatches:       111
Query Match:            41.34%         Indels:             3
DB:                                    Gaps:               3

US-09-308-397-2 (1-306) x US-09-974-300-1086 (1-960)
```

```
Qy    1  MetThrLysThrAlaPheLeuPheAlaGlyAlaGlnGlyAlaGlnTyrLeuGlyMetGlyArg   20
         ||||||||||||||||||||||||||||||||||||||||||||||
Db    1  ATGGGCCAAGATTGCTTTTCATTCCCGGCCAAGGTTCGCAGCATATCGCATGGGACAC       60

Qy   21  AspPheTyrAspGlnTyrProIleValLysGluThrIleAlaSerArgGlnValLeu       40
         :::  ::|:||::  :|||  ::::
Db   61  GAATTGTATGAAAAAGAACCGAATGCAAGATTTTTGAAGAAGCCGGATCAAACGCTT      120

Qy   41  GlyTyrAspLeuArgTyrLeuIle---AspThrGluGluLysLeuAsnGlnThrArg       59
         ||||:||::
Db  121  GAAACAAAACTGAGCACCCTCATGTTTGAAGGGGATCAAAAGGAACTGACGGTCACATAC   180

Qy   60  TyrThrGlnProAlaIleLeuAlaThrSerValAlaLeuTyrArgLeuGlnGluLys       79
         ||||:||||||||:||
Db  181  TTGAACTGCCCTGGCAAATCGTCATCTCGGGAACAGCTAAAGGGTGGAGCTCGCTTCA    240

Qy   80  GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyLeuTyrSerAlaLeuValAla       99
         |:||::||||:||
Db  241  GGCATTAAAGCCGACTATCGGCAGGTCACAGCCGCTTCGGCAATACAGCCATTGTGTCGCT   300

Qy  100  SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyMetGlyArg     119
         :::|||||||||||||||||||||:::|||
Db  301  GCCGGCGCCTTGTCGTTTAAAGATGCGGTTTATGCCGCAGAAGCGCGGCGAATTCATG      360

Qy  120  GluGluLeuGlnLysAlaGlyAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu   139
         ||:::||||||||||||:::
Db  361  AATGAAGCCGTGCCGGGCGGAGAAGGCGGCGGATGGCGGCGCCATTCTCGGCATGACGCAGCCAG             420

Qy  140  ValIleGluGluValAlaCysGlnLysLeuAlaSerGlyLeuGly---ValValThrProAlaAsn     158
         ||||:|||||||
Db  421  GCGCTGAAAGAAGTGACGGACAAAATTTCCGAAGAAGGAAACCTTGTTCAGCTCGCCAAT            480

Qy  159  TyrAsnThrProAlaGlnIleValIleAlaAlaGlyGluValAlaValAlaAlaArgAlaVal   178
         |||||:::
Db  481  TTGAACTGCCCTGGGCAAATCGTCATCTCGGGAACAGCTAAAGGCGTGGAGCTCGCTTCA    540

Qy  179  GluLeuLeuGlnAlaGlyAlaGlyLysArgLeuLeuProLeuLysSerGlyProPhe       198
         :::|||||||||:::
Db  541  GAGCTTGCGAAAGAAAAGGGCGCAAAACGCGGATTCCTCTGCGAAGTCAGCGGCCGTTC     600

Qy  199  HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaGluThrLeuAlaGlnValSer     218
         ||||:::
Db  601  CATTCTGAGCTGATGAAGCCGGAGCTAACGTAACGTTCGTGAAGTTCTTGATGCTGCACG   660

Qy  219  PheSerAspPheThrCysProLeuValGlyValGluAla--AlaValMetGlnLys        237
         ||||:::|:
Db  661  ATCAACGACGCCAGCCATTCGGCGTGTTCTCCAACGTAACGCCGACTTTGTAACGGATAAA   720

Qy  238  GluAspIleAlaGlnLeuLeuThrArgGlnValLeuGlyIleGluIleGlyTyrGluSer     257
         :::|||||||:::
Db  721  GACGACATTAAGAATAAAACTGATTGAACAGTCGTATTCCCGTACGCTTTGAAGAAACA    780

Qy  258  IleGlyValMetGlnHisGluAlaGlyTyrIleSerAsnPheIleGlyIleLeuGlyLysVal   277
         :::
Db  781  ATCAGCCGCCTGATTGACCGAAGGCGTCACGACCTTCATTGAAATCGGTCCCCGGAAAGGTT   840

Qy  278  LeuSerGlyPheValLysLysLeuAspAspThrAlaHisLeuAlaHisValGluAsp       296
         |||||||||:::
```

```
RESULT 8
US-09-974-300-1086
; Sequence 1086, Application US/09974300
; Patent No. US2002014672A1
; GENERAL INFORMATION:
;   APPLICANT: Berka, Randy M.
;   APPLICANT: Clausen, Ib Groch
;   TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;   TITLE OF INVENTION: Expression
;   FILE REFERENCE: 10085.500-US
;   CURRENT APPLICATION NUMBER: US/09/974,300
;   CURRENT FILING DATE: 2001-10-05
;   PRIOR APPLICATION NUMBER: 09/680,598
;   PRIOR FILING DATE: 2000-10-06
;   PRIOR APPLICATION NUMBER: 60/279,526
;   PRIOR FILING DATE: 2001-03-27
;   NUMBER OF SEQ ID NOS: 8481
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1086
```

```
Db   61  GAATTGTATGAAAAAGAACCGAATGCAAGATTTTTGAAGAAGCCGGATCAAACGCTT      120

Qy   41  GlyTyrAspLeuArgTyrLeuIle---AspThrGluGluLysLeuAsnGlnThrArg       59

Db  121  GAAACAAAACTGAGCACCCTCATGTTTGAAGGGGATCAAAAGGAACTGACGGTCACATAC   180

Qy   60  TyrThrGlnProAlaIleLeuAlaThrSerValAlaLeuTyrArgLeuGlnGluLys       79

Db  181  AACGGCGCCAGCCAGCCTTTTACGCGGCCATCGAGCGCGCTTGAAAAACTGAAGGAATAC   240

Qy   80  GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyLeuTyrSerAlaLeuValAla       99

Db  241  GGCATTAAAGCCGACTATCGGCAGGTCACAGCCGCTTCGGCAATACAGCCATTGGTCGCT   300

Qy  100  SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyMetGlyAlaTyrMet  119

Db  301  GCCGGCGCCTTGTCGTTTAAAGATGCGGTTTATGCCGCAGAAGCGCGGCGAATTCATG      360

Qy  120  GluGluLeuAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGlu    139

Db  361  AATGAAGCCGTGCCGGGCGGAGAAGAAAAGGGCGCAAAACGCGGCCATTCTCGGCATGACGGCCAG   420

Qy  140  ValIleGluGluValAlaCysGlnLysLeuAlaSerGluLeuGly---ValValThrProAlaAsn  158

Db  421  GCGCTGAAAGAAGTGACGGACAAAATTTCCGAAGAAGGAAACCTTGTTCAGCTCGCCAAT    480

Qy  159  TyrAsnThrProAlaGlnIleValIleAlaAlaGlyGluValAlaValAlaAspArgAlaVal  178

Db  481  TTGAACTGCCCTGGGCAAATCGTCATCTCGGGAACAGCTAAAGGGTGGAGCTCGCTTCA    540

Qy  179  GluLeuLeuGlnGluGlnAlaGlyAlaGlyLysArgLeuLeuProLeuLysSerGlyProPhe   198

Db  541  GAGCTTGCGAAAGAAAAGGGGCGCAAAACGCGGATTCCTCTGCGAAGTCAGCGGCGGCCGTTC    600

Qy  199  HisThrAlaLeuLeuGluGluLysGlnLysLeuLeuThrGluLeuAlaGlnValSer       218

Db  601  CATTCTGAGCTGATGAAGCCGGTCGTCCACACGTAACGTTCGTGAAGTTCTTGATGCTGCACG  660

Qy  219  PheSerAspPheThrCysProLeuValGlyValAlaAsnThrGluAla--AlaValMetGlnLys  237

Db  661  ATCAACGACGCCAGCCATTCCGGCGTGTTCTCCAACGTAACGGCCGACTTTGTAACGGATAAA  720

Qy  238  GluAspIleAlaGlnLeuLeuThrArgGlnLeuLeuGluLeuThrGluGlyIleGluIleGlyTyrGlySer  257

Db  721  GACGACATTAAGAATAAAACTGATTGAACAGTCGTATTCCCGTACGCTTTGAAGAAGACA   780

Qy  258  IleGlyValMetGlnHisGluAlaGlyTyrIleSerAsnPheIleGlyIleLeuGlyLysVal  277

Db  781  ATCAGCCGCCTGATTGACGAAGGCGTCACGACCTTCATTGAAATCGGTCCCCGGAAAGGTT  840

Qy  278  LeuSerGlyPheValLysLysLeuAspGlnThrAlaHisLeuAlaHisValGluAsp      296

Db  841  TTGTCAGGGCTTGTGAAGAAGTGAACCGCAGAGTCAAAACGATTGCTGTATCAGAC      897
```

RESULT 9
US-09-815-242-4405
; Sequence 4405, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR APPLICATION FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR APPLICATION FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR APPLICATION FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR APPLICATION FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR APPLICATION FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR APPLICATION FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR APPLICATION FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4405
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4405

Alignment Scores:
Pred. No.:
Score: 1.67e-58   Length: 906
                      556.50   Matches: 122
Percent Similarity: 59.06%   Conservative: 54
Best Local Similarity:40.94%   Mismatches: 117
Query Match: 36.66%   Indels: 5
DB: 10   Gaps: 4

US-09-308-397-2 (1-306) x US-09-815-242-4405 (1-906)

Qy   1 MetThrLysThrGlnPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg   20
         :::|||||||||||||||||||||||| ::|:|||||||
Db   1 ATGAGTAAAACAGCAATTATTTTTCCGGGACAAGGTGCCCAAAAGTTGGTATGGCACAA   60

Qy   21 AspPheTyrAspPheValProIleValLysGluThrIleAspArgAlaSerGlnValLeu   40
         |||| :::::|||||||
Db   61 GATTATATAACAACAACGATCAAGCAACTGAAATTTAACTTCAGCAGACAAACACGTTA   120

Qy   41 GlyTyrAspLeuArgTyrLeuIleAspThrGluAsp--LysLeuAsnGlnThrArg   59
         :::|::::: ||| ::|||
Db 121 GACTTTGATATTTAGAGACAATGTTTACTGACGAAGACGGCAAATTGGGTGAAACTGAA   180

Qy   60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys   79
         |||:::   ::|:||
Db 181 AACACGCAACCAGCTTTATTGACGACAGCCGCGTTAGCGGCGCGTA------AAA   234

Qy   80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla   99
         ||||| |||| :::|||||
Db 235 AATTTGAATCCCGATTTTACTATGGGCATAGTTTAGGTGAATATTCAAGTTAGTTGCA   294

Qy 100 SerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysAlaArgGlyAlaTyrMet   119

```
;  SEQ ID NO 8032
;  LENGTH: 936
;  TYPE: DNA
;  ORGANISM: Staphylococcus aureus
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(936)
US-09-815-242-8032

Alignment Scores:
Pred. No.:              1.75e-58        Length:        936
Score:                  556.50          Matches:       122
Percent Similarity:     59.06%          Conservative:   54
Best Local Similarity:  40.94%          Mismatches:    117
Query Match:            36.66%          Indels:          5
DB:                                     Gaps:            4

US-09-308-397-2 (1-306) x US-09-815-242-8032 (1-936)

Qy    1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg 20
          |||:::|        :::|||        |   |::|   |:|||
Db   10 ATGAGTAAAACAGCAATTATTTTTCCGGGACAAGGTGCCCAAAAAGTTGGGATGGCACAA 69

Qy   21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu 40
          || |||    |   ||||||            ||   |  ||| |||
Db   70 GATTTATATAACACAACGATCAAGCAACTGAAATTTTAACTTCAGCAGCAAACACGTTA 129

Qy   41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGluAsp---LysLeuAsnGlnThrArg 59
          ::|::|        |||  || |||       |||
Db  130 GACTTTGATATTTTAGAACAATGTTTACTGACGAAGACGGCAAATTGGGTGAAACTGAA 189

Qy   60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys 79
          |||         ::|        |      |:     :::
Db  190 AACACGCCACCAGCTTTATTGACGCATAGCTCGGCGTTATTAGCCGGGCTA-----AAA 243

Qy   80 GlyTyrGlnProAspMetValAlaGlyGlyLeuSerLeuGlyGluTyrSerAlaLeuVal 99
          :::|||    ||| ||||       ||:|:  |:
Db  244 AATTTGAATCCCGATTTTACTATGGGCATAGTTTAGGTGAATATTCAAGTTTAGTTGCA 303

Qy  100 SerGlyAlaLeuAspPheGluAsnGlyLysGlnLys---AlaSerGluLeuGlyValCys 119
          :::       |||        |::||       ||
Db  304 GCTGACGATTATCATTTGAAGAGTGCAGTTAAAATTGTTAGAAAACGTGGTCAATTAATG 363

Qy  120 GlnLysGlnLeuAlaLeuAsnTyrLysMetValAlaValLeuAsnThrProValGlu 139
          :||||||      :::|                 |||
Db  364 GCGCAAGCATTTCCTACTGGTGTGAGGAAGCATGGCGTGGTTGGGCTTAGATTTGAT 423

Qy  140 ValIleGluGluLysGlnLys---AlaSerGluLeuGlyValValThrProAlaAsn 158
          |||  |||                :::
Db  424 AAAGTCGATGAAATTGTAAGTCAGAAACGTCATCCTGATGACAAAATAATTGAACCAGCAAAC 483

Qy  159 TyrAsnThrProAlaGlnIleValIleAlaGlyGluValAlaValAlaAspArgAlaVal 178
          |||||:::                |||                ::|  |||
Db  484 ATTAATTGCCCAGGTCAAATTGTTGTTCAGGTCACACAAAAGCTTAATTGATGAGCTAGTA 543

Qy  179 GluLeuLeuGlyGluAlaGlyLysLeuAlaLysArgLeuIleProLeuLysValGlyProPhe 198
          |||||:|||     |||:|
Db  544 GAAAAAGTAAATCATTAGGTGCAAAACGTCATGCCTTTAGCAGTAATCTGGCCCATTC 603

Qy  199 HisThrAlaLeuLeuGluProAlaSerGlnLysLeuAlaLeuAlaGluThrLeuGlnIleValSer 218
          :::                :::   :::
Db  604 CATTCATCGCTATGAAAGTGCCTAATGAAGATTTTTCAAGTTATATTAATCAATTAGAA 663

Qy  219 PheSerAspPheThrCysProLeuValGlyValGlyValAsnThrGluAla---AlaValMetGlnLys 237
          ::|    |||    :||||||  :::|||  ||
Db  664 TGGCATGATGCTAAGTTCCTGTAGTTCAAAAATGTAAATGCCCAAGTGAAACTGACAAA 723

Qy  238 GluAspIleAlaGlnIleLeuLeuThrArgGlnValLysGluProValLysProPheTyrGluSer 257
          :||
Db  724 GAAGTAATTAAAATCTAATATGTCAAACAATTATATTCACCAATTCATTAACTCA 783

Qy  258 IleGlyValMetGlnGluGluAlaGlyLyIleSerAsnPheIleGluIleGlyProGlyLysVal 277
          ::: ::|
Db
```

```
Db  784 ACAGAATGGCTAATAGACCAAGGTGTTGATCATTTATTGAAATGGTCTGGAAAAGTT 843
          ||||||||                       :::|||
Qy  278 LeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGlu 295
          :::|||                |||:::
Db  844 TTGTCTGGCTTAATTAAAAAAATAAATAGAGATGTAAGTAACATCAATTCAA 897
```

```
RESULT 11
US-09-815-242-7796
;  Sequence 7796, Application US/09815242
;  Patent No. US20020061569A1
;  GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT: Ohlsen, Kari L.
;  APPLICANT: Zyskind, Judith W.
;  APPLICANT: Wall, Daniel
;  APPLICANT: Trawick, John D.
;  APPLICANT: Carr, Grant J.
;  APPLICANT: Yamamoto, Robert T.
;  APPLICANT: Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION: Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE:  2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE:  2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE:  2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE:  2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE:  2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE:  2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 7796
;  LENGTH: 939
;  TYPE: DNA
;  ORGANISM: Pseudomonas aeruginosa
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(939)
US-09-815-242-7796

Alignment Scores:
Pred. No.:              1.76e-58        Length:        939
Score:                  556.50          Matches:       126
Percent Similarity:     62.54%          Conservative:   51
Best Local Similarity:  44.52%          Mismatches:    103
Query Match:            36.66%          Indels:          3
DB:                                     Gaps:            3

US-09-308-397-2 (1-306) x US-09-815-242-7796 (1-939)

Qy    5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
          :::
Db   16 GCATTCGTCTTCCCTGGCCAGGGTTCGCAATCCCTGGCCGAGCTGGGCGGCC 75

Qy   25 GlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAspLeu 44
          |||
Db   76 CAGCAGGGCCGTGGTGCGGATACCTTCGCCGAGGCCTCCGAGGGCGTCGGTTACGACCTT 135

Qy   45 ArgTyrLeuIleAspThr---GluGluAspLysLeuAsnGlnThrArgTyrThrGlnPro 63
          ::|
Db  136 TGGGCGCTGGTCGACAAGATGGCCTGAAGATGCCGAACCAGACCAGCCG 195

Qy   64 AlaIleLeuAlaThrSerValAlaIleTyrArgLeu--LeuGlnGluLysGlyTyrGln 82
          |||||           :::|||
Db
```

ATTORNEY/AGENT INFORMATION:
  NAME: Benson, Bob
  REGISTRATION NUMBER: 30,446
  REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (301) 309-8504
  TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
  LENGTH: 6251 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
US-08-781-986A-25

Alignment Scores:
  Pred. No.:   Length: 6251
  Score: 1.54e-56   Matches: 121
  Percent Similarity: 550.50   Conservative: 53
  Best Local Similarity: 58.39%   Mismatches: 119
  Query Match: 40.60%   Indels: 5
  DB: 36.26%   Gaps: 4
                         7

US-09-308-397-2 (1-306) x US-08-781-986A-25 (1-6251)

Qy   1 MetThrLysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlnGlyMetGlyArg   20
           |||:::||||:::   :::|:|| ::|||||||:
Db 3386 ATGAGTAAAACGACAATTATTTTCCGGGACAAGGTGCCCAAAAAGTGGTATGGCGCCAA   3445

Qy   21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu   40
           :|:|||| :|||||::|||| ||||:::|||
Db 3446 GATTTGTTTAACACACATGATCAAGCAACTGAAATTTAACTTCAGCAGCGAACACATTA   3505

Qy   41 GlyTyrAsp---LeuArgTyrLeuIleAspThrGluAspLysLeuAsnGlnThrArg   59
           :::|||| |||:||| |:||||
Db 3506 GACTTTGATATTTAGAGACAATGTTACTGATGAAGAAGGTAAATTGGGTGAAAACTGAA   3565

Qy   60 TyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLys   79
           ||| ::: ||| ||
Db 3566 AACACACACCAGCTTATTGACGCACATAGTTCGCCATTATTAGCAGCGCTA------AAA   3619

Qy   80 GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAla   99
           :::||| |:|| :|| ||:::
Db 3620 AATTTGAATCCTGATTTTACATGGGCATAGTTTAGGTGAATATTCAAGTTTAGTTGCA   3679

Qy   100 SerGlyAlaLeuAspPheGluAspAlaLeuLysArgLeuIleLeuLysValSerGlyProPhe 119
           ::: |||||||:::|| ||
Db 3680 GCTGACGTATTATCATTTGAAGATGCAGTTAAAATTGTTAGAAAAACGTGGTCAATTAATG   3739

Qy   120 GluGluAlaAlaProAlaAspSerGlyLysMetValAlaLeuAsnThrProValGlu   139
           |||||||||||:|: |||
Db 3740 GGGCAAGCGATTTCCTACTGGTGTAGGAGAGCATGGCTGCAGTATTGGGATTAGATTTTGAT   3799

Qy   140 ValIleGluGlyAlaCysGlnLys---AlaSerGluLysGlyValValThrProAlaAsn   158
           |||||||||||||| ::||
Db 3800 AAAGTCGATGAAATTGTAAGTCATTATCAACTGATGACAAAATAATTGAACCAGCAAAC   3859

Qy   159 TyrAsnThrProAlaGlnIleValIleGluAlaGlyValValAlaValAlaAspArgAlaVal   178
           ::::: ||| ||||
Db 3860 ATTAATTGCCCAGGTCAAATTGTGTTCACCAAGCTTAATTGATGAGCTAGTA   3919

Qy   179 GluLeuLeuGlnGluAlaGlyIleAlaLeuArgLeuIleLeuLysValSerGlyProPhe   198
           ::| :: ||
Db 3920 GAAAAAGTAAATCATTAGGTGCAAAACGTGTCATGCCTTTAGCAGTATCTGACCATTC   3979

Qy   199 HisThrAlaLeuLeuGluProAlaSerGlnLysLeuLysLeuGluThrGlnValSer   218
           |||||| :::|| |||
Db 3980 CATTCATCGCTAATGAAGATGATTGAAGAAGATTTTCAGTTACATTAATCAATTTGAA   4039

Qy   219 PheSerAspPheThrCysProLeuValGlyAsnThrGluAla--AlaValMetGlnLys   237
           :::||||::|:   :::|| |||
Db 4040 TGGCGTGATGCTAAGTTTCCTGTAGTTCAAAAATGTAAATGCCAAGGTGAAACTGACAAA   4099

Db   196 GCCATCCTTACGGTTTCGATCGCGCTCTGCGCCTCTGGCTGGCCGAGGGCGGTGCGCGGC   255

Qy   83 ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyAla   102
           ||| :::| ::
Db   256 CCGGGCGTTCGTCGCCGGGCACAGCCTGGGCGAATATTCCGCGTCCGGCCGAAAGC   315

Qy   103 LeuAspPheGluAspAlaLeuLysArgLeuIleLeuLysValSerGlyProPheAla   122
           |||||||:::| ||||
Db   316 CTGGCGTTCGTCGCCGATGCGGTTCAAGCTGGTCGAGCGGTAGGGGCCAACTGATGCGAGCAGGCG   375

Qy   123 AlaProAlaAspSerGlyLysMetValAlaLeuAsnThrProValGluValIleGlu   142
           ||||:::||
Db   376 GTTCCGGCGCGGGCGGCAGGCCGCGATCCTTGGGCTGGAAGAGACCCCGATGTATTG   435

Qy   143 GluAlaCysGlnLysAlaSerGlyIleLeuGlyValValThrProAlaAsnTyrAsnThrPro   162
           ::::: |:::
Db   436 GCGGGCCTGTGCCGAGGCGGCGCCCAGGGCGAGCTGGTCAGCGCGGTCAACTTCAACGCGCGC   495

Qy   163 AlaGlnIleValIleGluAlaGlyGluValValAlaValAlaAspArgAlaValGluLeuLeuGln   182
           |||:::|||:::||||
Db   496 GCGCAGGTAGTGATCGCCGGTGCCCGGCGTGCCGCGGTGCCATCGAGGCATGCAAG   555

Qy   183 GluLeuAlaGlyLeuLysArgLeuIleLeuProLeuLysSerGlyProPheHisThrAlaLeu   202
           |||| :::
Db   556 GCACGCGGCGCCCAAGGCGCGCGGGTGGCGTTGCCAGTCAGCGCGTCCGCCATTGCGAACTG   615

Qy   203 LeuGluProAlaSerGlnLysGlnIleLysLeuLeuAlaGlyLysThrLeuGlnIleValSerPheSerAspPhe   222
           ::   ::|
Db   616 ATGCGTCCGCGCCGAGCAGTTCCCGCCTCGGTCGAAAGCCCTCGAGCCTGCAGGCGCCG   675

Qy   223 ThrCysProLeuValGlyTyrAsnThrGluAlaAlaValMetGlnLysGlnGlyAsp---IleAla   241
           ||
Db   676 AAGATTTCGCTGGTCAGAACGTCAGGAACGTCCGGCGACCTGATCTGATACGCTGCGC   735

Qy   242 GlnLeuLeuThrArgGlnValLysProValArgPheTyrGluSerIleGlyValMet   261
           |||:::|||| |:::   :::
Db   736 CGGACCCTGCGACACAGTCGTACAGCCCCGTGTTCGCTGGTGGAGCATCCAGCTGCTG   795

Qy   262 GlnGluAlaGlyIleLeuSerAsnPheIleGlyIleGlyLeuGlyProGlyLysValLeuuSerGlyPhe   281
           ||| :::|||
Db   796 GCGGAAAAGGGCGTCACCGAGCTGGTCGGAGTGCGGCCGGGCCAGGTCCTGCAGGCCTC   855

Qy   282 ValLysLys   284
           :::||||:
Db   856 AACAGGCGC   864

RESULT 12
US-08-781-986A-25
; Sequence 25, Application US/08781986A
; Publication No. US20030054436A1
GENERAL INFORMATION:
  APPLICANT: Charles Kunsch
  TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
  NUMBER OF SEQUENCES: 5255
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Human Genome Sciences, Inc.
  STREET: 9410 Key West Avenue
  CITY: Rockville
  STATE: Maryland
  COUNTRY: USA
  ZIP: 20850
COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
  COMPUTER: HP Vectra 486/33
  OPERATING SYSTEM: MSDOS version 6.2
  SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/781,986A
  FILING DATE:
  CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER:
  FILING DATE:

```
Qy   238  GluAspIleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSer  257
          ||  ::|||          ::|||||          :::|:|          ||||
Db  4100  GAAGTAATTAAATCTAATATGGTCAAGCAATTATTCACCAGTACAATTCATTAACTCA  4159

Qy   258  IleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGlyIleLeuGlyProGlyLysVal  277
          ||||                    |||:|||          |||||
Db  4160  ACAGAAATGGCTATAGACCAAGTGTTGATCATTTATTGAAATTGGTCCTGGAAAAGTT  4219

Qy   278  LeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGlu  295
          |||||          :::|||||:::::|
Db  4220  TTATCTGGCTTAATTAAAAAAATAAATAGAGATGTTAACATCAATTCAA  4273
```

RESULT 13
US-09-070-927A-458

; Sequence 458, Application US/09070927A
; Patent No. US2002012011GA1
; GENERAL INFORMATION:
;   APPLICANT: Charles A. Kunsch
;   Patrick J. Dillon
;   Steven Barash
;   TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
;   NUMBER OF SEQUENCES: 982
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/070,927A
;   FILING DATE: 04-May-2000
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/046,655
;   FILING DATE: 1997-05-16
;   APPLICATION NUMBER: 60/044,031
;   FILING DATE: 1997-05-06
;   APPLICATION NUMBER: 60/066,009
;   FILING DATE: 1997-11-14
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Kenley K. Hoover
;   REGISTRATION NUMBER: 40,302
;   REFERENCE/DOCKET NUMBER: PB369
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 458:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6021 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 458:
US-09-070-927A-458

Alignment Scores:
Pred. No.:   Length: 6021
Score: 9.59e-41   Matches: 84
Percent Similarity: 421.00   Conservative: 23
Best Local Similarity: 70.86%   Mismatches: 44
Query Match: 55.63%   Indels: 0
DB: 27.73%   Gaps: 0
                    10

US-09-308-397-2 (1-306) x US-09-070-927A-458 (1-6021)

```
Qy   154  ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleValAlaGlyGluValValAla  173
          |||          ||||||||||          ||||||||
```

```
Qy   173  GTGGCTCCAGCAAATTATWATWCMCMCAMCAAATCGTGATTGGTGGTGAGGTTGCTGCT  232

Db   174  ValAspArgAlaValGluLeuGlnGluLeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLys  193
          ||||||          :::::                    |||||
          GTTGATCAAGCGATGACACTTCTCAAAGAGCTGGTGTGAAGCGAATGATTCCGTTAAAT  292

Db   233

Qy   194  ValSerGlyProPheHisThrAlaLeuGluLeuGluProAlaSerGlnLysLeuAlaGluThr  213
          |||||          :::                    |||
Db   293  GTGAGTGGCCCTTTCCATACGGCGCTGTTACAACCAGCATCAAAAAAATTGGCTCAGGAT  352

Qy   214  LeuAlaGlnIleValSerPheSerAspPheThrCysProLeuValGlyValAsnThrGluAlaAla  233
          |||:::                    :::
Db   353  TTAGCAAAATTGAACTTTCAAACGATGCAAATTCCTGTCATTAGTAATACGACTGCCGAA  412

Qy   234  ValMetGlnLysGluAspIleAlaGlnLeuLeuThrArgGlnIleLysGluProValArg  253
          ::||||  ::|||||          ::|:|||
Db   413  ATTATGCCCAGAGGCAATTCAAGCGTTATTGGAAAAGCAAGTCATGTCGCGTAACGT  472

Qy   254  PheTyrGluSerIleGlyValMetGlnGluAlaGlyIleSerAsnPheIleGlyIleGly  273
          |||                    :::
Db   473  TTTGAAGACAGTATCGAAACGATGAAGGCTATGAACGTAGGAACGATGATTGAGTTGGT  532

Qy   274  ProGlyLysValLeuSerGlyPheValLysLysIleAspGlnThrAlaHisLeuAlaHis  293
          |||:::                    |||:::
Db   533  CCAGGGAAAACATTAACTGGTTTTGTTAAAAAAATTGACAAAACAATTGAAATGACCGT  592

Qy   294  ValGluAspGlnAlaSerLeuValAlaLeuLeu  304
          :::|||
Db   593  GTGGAAGATGTTGCACATTAACAGAAACGTTA  625
```

RESULT 14
US-09-815-242-7193

; Sequence 7193, Application US/0981524
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari L.
;   APPLICANT: Zyskind, Judith W.
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John D.
;   APPLICANT: Carr, Grant J.
;   APPLICANT: Yamamoto, Robert T.
;   APPLICANT: Xu, H. Howard
;   TITLE OF INVENTION: Identification of Essential Genes in
;   TITLE OF INVENTION: Prokaryotes
;   FILE REFERENCE: ELITRA.011A
;   CURRENT APPLICATION NUMBER: US/09/815,242
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: 60/191,078
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: 60/206,848
;   PRIOR FILING DATE: 2000-05-23
;   PRIOR APPLICATION NUMBER: 60/207,727
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: 60/242,578
;   PRIOR FILING DATE: 2000-10-23
;   PRIOR APPLICATION NUMBER: 60/253,625
;   PRIOR FILING DATE: 2000-11-27
;   PRIOR APPLICATION NUMBER: 60/257,931
;   PRIOR FILING DATE: 2000-12-22
;   PRIOR APPLICATION NUMBER: 60/269,308
;   PRIOR FILING DATE: 2001-02-16
;   NUMBER OF SEQ ID NOS: 14110
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7193
;   LENGTH: 930
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(930)
US-09-815-242-7193

```
Alignment Scores:
Pred. No.:           2.18e-41      Length:           930
Score:               416.50        Matches:          106
Percent Similarity:  51.80%        Conservative:     52
Best Local Similarity: 34.75%      Mismatches:       138
Query Match:         27.44%        Indels:           9
DB:                                Gaps:             5

US-09-308-397-2 (1-306) x US-09-815-242-7193 (1-930)

Qy      5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp  24
            |||      ||||||| :: :||:|||:
Db     10 GCGGTTATTATTTCCAGGCCAAGGCTCGCAATGTGTAGGAATGGGAAATCATTCTATGAG  69

Qy     25 GlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAspLeu  44
            : :: |||
Db     70 AGCCACCACTCTAGCTCAAAGAGAATTGTTGAAAGGGCTTCAACCGCACTTAAAGTGGAT  129

Qy     45 ArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGlnProAla  64
            :::     ||||| |||      |||||||||
Db    130 AAAAAAACCGTTTTGAAGAAAATGAGTTTTAAAAGAAGAGCGCCTTACACCCAGCCTGCC  189

Qy     65 IleLeuAlaThrSerValAlaIleIleTyrArgLeuLeuGlnGluLys---------GlyTyr  81
            ||| |    :::::
Db    190 ATTTATTTAGTGAGCTATATCGCTTACCAATTGCTCAACAAGCAAGTAAATGGGGGGGTA  249

Qy     82 GlnProAspMetValAlaGlyMetAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGly  101
            ::::: ||||||||
Db    250 AAACCCGTTTTGCTTTGGCATTCGCTCGGCGAAGTGAGCGCGGTGTCTTTGAGTGGG  309

Qy    102 AlaLeuAspPheGluAspAlaValAlaLeuValAlaLeuValAlaArgGlyAlaTyrMetGluGlu  121
            ||| ::: ||:||:|||||||
Db    310 GCGTTAGATTTTTGAAAAAGCCCTTAAACTCACGCACCAAAATGATGCAAGAA  369

Qy    122 AlaAlaProAlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGluValIle  141
            :::
Db    370 GCGGTCGCGGATAAAGACGCTTCCATGATGGTCGTTTTGGCGTTTCTGAAGAAAGCCTT  429

Qy    142 GluGluAlaCysGlnLysAlaSerGlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThr  161
            ||| |||
Db    430 TTGAGTTTGTGTCAAAGAACCAAAAT--------GTGTGGTGCGCGAATTTCAATGGC  480

Qy    162 ProAlaGlnIleValIleIleAlaGlyValGluValAlaValAlaAspArgAlaValGluLeuLeu  181
            ||||| :::||:|||||
Db    481 GGCATGCAAGTGGTTTAGCCGGGATTAAAGACGATTTGAAAGCCCTAGAGCCGACTTTA  540

Qy    182 GlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSerGlyProPheHisThrAla  201
            ||| :::||
Db    541 AAGGAAAATGGGGGCTAAAAGAGTGGTTTTTTAGAAATGAGCGTGGCGGCCATTGCCCT  600

Qy    202 LeuLeuGluProAlaSerGlnLysLeuGluThrLeuAlaGlnInValSerPheSerAsp  221
            : :::
Db    601 TTTTTAGAGCTTATGACTTTTAAATTCCAGGAATTGCTAGAAAAAGCTGAAAGATAAA  660

Qy    222 PheThrCysProLeuValGlyAsn---ThrGluAlaAlaValMetGlnLysGluGluAspIle  240
            :::||:|||
Db    661 TTCCATTTGAAATCATCTCCAATGCGACTAATGAACGCTATCACACAAAGCAAAGCC  720

Qy    241 AlaGlnLeuThrArgGlnInValLysGluProValArgProTyrPheLysValLeuSerGly  260
            :::||: ||||: ||:
Db    721 GTTGAACTATTGAGCTTGCCAACTCACTCCAGCCGGTGCCGTTATCAAGACTGCGTGAAATCC  780

Qy    261 MetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGly  280
            ::: ||:
Db    781 AACAATGACCGA---GTGGATGGTCTTTTTTTGAATTAGGCTGTGGAGTGTGTTAAAGGGGG  837

Qy    281 PheValLysLysIleAspGlnThrAlaHisLeuAlaHisValGluAspGlnAlaSerLeu  300
            :::: |||
Db    838 CTTAACAAGCCGATTAAGCAACAACAAACCCCACCATAAGC---GTGGGGGATAATAAAGGGCTT  894

Qy    301 ValAlaLeuLeuGlu  305
            ::::|||
Db    895 GATGAAGGCATTGAG  909
```

```
RESULT 15
US-09-974-300-5552
; Sequence 5552, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;   APPLICANT: Berka, Randy M.
;   APPLICANT: Clausen, Ib Groth
;   TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;   TITLE OF INVENTION: Expression
;   FILE REFERENCE: 10085.500-US
;   CURRENT APPLICATION NUMBER: US/09/974,300
;   CURRENT FILING DATE:  2001-10-05
;   PRIOR APPLICATION NUMBER: 09/680,598
;   PRIOR FILING DATE: 2000-10-06
;   PRIOR APPLICATION NUMBER: 60/279,526
;   PRIOR FILING DATE: 2001-03-27
;   NUMBER OF SEQ ID NOS: 8481
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 5552
;     LENGTH: 592
;     TYPE: DNA
;     ORGANISM: Bacillus clausii
US-09-974-300-5552

Alignment Scores:
Pred. No.:            1.29e-41      Length:           592
Score:                416.00        Matches:          93
Percent Similarity:   53.59%        Conservative:     31
Best Local Similarity: 47.69%       Mismatches:       69
Query Match:          27.40%        Indels:           2
DB:                                 Gaps:             2

US-09-308-397-2 (1-306) x US-09-974-300-5552 (1-592)

Qy      3 LysThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPhe  22
            ||||| |:::||:|||: |||||||:
Db      7 AAAGTTGCCTTTTTATTTCCTGGACAAGGCTCACAAAAAGTTGGCATGGGAGCAACGCTT  66

Qy     23 TyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyr  42
            ::: ||||||: :::||:|||
Db     67 TTTGCTGAAGATTGATGGCAAGGAGTGGCTTGAGCGGGTCGACCAAGGCGCTTGGTTTT  126

Qy     43 AspLeuArgTyrLeuIleAsp---ThrGluGluAspLysLeuAsnGlnThrArgTyrThr  61
            :::|||||||
Db    127 CCTCTTTCAGAGATCATCGAACATGGACGAGAGACCAAGCTCGAAACAAACGGCTTATGCC  186

Qy     62 GlnProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGlyTyr  81
            ::: ||||||
Db    187 CAACCTGCTTTAGTGACGATGAGCACTGCAGTCCTCCAATGTTCCGCGAGCTGACATT  246

Qy     82 GlnProAspMetValAlaGlyMetAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGly  101
            :::||||||
Db    247 CAAGCACGATTTTGTTGCGAGCCATAGCCTAGGGAGAGTTAGTGCCATGTGAAG  306

Qy    102 AlaLeuAspPheGluAspAlaValAlaLeuValAlaLeuValAlaLysArgGlyAlaTyrMetGluGlu  121
            :::||: |||||||
Db    307 TCCCTTACATTTGAAGATGCGTCACACTTGTTCATCAACGGGGACACTCATGAGGAA  366

Qy    122 AlaAlaProAlaAspSerGlyLysMetValAlaValAlaLeuAsnThrProValGluValIle  141
            ::: ||:
Db    367 GCTGTGCCACAAGGACAAGGGCGATGGCGGTACTGGCTTAAATAAAGAGAGCTT  426

Qy    142 GluGlu---AlaCysGlnLysAlaSerGluLysGlyValValThrProAlaAsnTyrAsn  160
            ::: :::
Db    427 GAGGAAGTGGCTTCCGAAATTGCAGGCGGACGGCGAAGTTGCTGAACTTGAAT  486

Qy    161 ThrProAlaGlnIleValIleIleAlaGlyValGluValAlaValAlaAspArgAlaValGluLeu  180
            ||| |||
Db    487 TGTTCAGGGCAAATTGTCCTGTCAGGGACTCGGAAAGGAATTGAACAAGCAGCAGTATTA  546

Qy    181 LeuGlnGluAlaGlyAlaLysArgLeuIleProLeuLysValSer  195
            :::::||:|||||
```

Db   547 GCAAAACAAAAAGGGGCCAAGCCGTGTCTTGCCACTTGCCGTCAGC 591

Search completed: June 11, 2003, 22:27:08
Job time : 774 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: June 11, 2003 - 20:10:52 ; Search time 1430 Seconds
(without alignments)
3465.608 Million cell updates/sec

Title: US-09-308-397-2
Perfect score: 1518
Sequence: 1 MTKTAFLFAGQGAQYLGMGR.........QTAHLAHVEDQASLVALLEK 306

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 33308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09308397/runat_06062003_112349_28885/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINNLEN=0 -MAXNLEN=2000000000
-USER=US09308397@CGN_1_1_1525 @runat_06062003_112349_28885 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=100 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_eston:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		327	21.5	787	12	BG584476	BG584476 EST486236
2	c	324.5	21.4	511	14	BQ818656	BQ818656 1030072D0
3	c	314.5	20.7	544	9	AU069821	AU069821 AU069821
4		296.5	19.5	647	9	AJ500905	AJ500905 AJ500905
5		293.5	19.3	690	9	AJ498681	AJ498681 AJ498681
6		291	19.2	647	12	BG851744	BG851744 1024032B1
7		290	19.1	585	10	BE404963	BE404963 WHE1207_B
8		289.5	19.1	491	12	BF512296	BF512296 UI-H-BW1_
9		278.5	18.3	648	10	BE440632	BE440632 sp50b03·y
10		278	18.3	539	13	BJ462718	BJ462718 BJ462718
11		274.5	18.1	850	12	BG444706	BG444706 GA_Ea002
12		274	18.1	513	13	BJ462722	BJ462722 BJ462722
13	c	270	17.8	892	11	AY108033	AY108033 Zea mays
14		270	17.8	1237	17	AF029428	AF029428 AF029428
15		269	17.7	599	10	BE237947	BE237947 894039B11
16		269	17.7	708	12	BG441502	BG441502 GA_Ea001
17		267	17.6	659	12	BE725078	BE725078 89408OH05
18		266	17.5	684	12	BG441246	BG441246 GA_Ea001
19		265.5	17.5	809	12	BG418889	BG418889 HVSMEk002
20		264	17.4	658	12	BG591410	BG591410 EST499252
21		264	17.4	744	12	BG597049	BG597049 EST495727
22		263.5	17.4	827	13	BM359182	BM359182 GA_Ea001
23		263	17.3	648	9	AJ500837	AJ500837 AJ500837
24		260.5	17.3	602	13	BM525587	BM525587 sa129g11.
25		258	17.0	587	13	BM524467	BM524467 sa116a10.
26		257	16.9	742	13	BI934864	BI934864 EST554753
27		256	16.9	582	10	BE443678	BE443678 WHE1121_E
28		253	16.7	816	12	BF630807	BF630807 HVSMEb001
29		252.5	16.6	585	13	BJ314706	BJ314706 BJ314706
30		252	16.6	1104	12	BF795116	BF795116 601592648
31		252	16.6	1106	12	BF789052	BF789052 602678651
32		250.5	16.5	613	10	AV926272	AV926272 AV926272
33		250	16.4	582	10	AW617993	AW617993 EST14067
34		248	16.3	668	12	BG859575	BG859575 1024064B1
35		247.5	16.3	604	10	AV974832	AV974832 AV974832
36		246	16.2	755	13	BI919675	BI919675 EST539610
37		246	16.2	788	13	BI436174	BI436174 EST538935
38		244	16.1	628	12	BG846299	BG846299 1024013B0
39		243	16.0	577	10	AV924867	AV924867 AV924867
40		243	16.0	601	13	BI064433	BI064433 pgfin.pk0
41		242	15.9	642	14	BQ462913	BQ462913 HI02G11r
42		241	15.9	796	12	BG699020	BG699020 602678651
43		239.5	15.8	534	9	AJ499573	AJ499573 AJ499573
44		239	15.7	575	13	BI500434	BI500434 rs72e12·y
45		239	15.7	699	9	AL508847	AL508847 AL508847

ALIGNMENTS

RESULT 1
BG584476
LOCUS   BG584476   787 bp   mRNA   linear   EST 11-APR-2001
DEFINITION EST486236 MHAM Medicago truncatula/Glomus versiforme mixed EST
            library cDNA clone pMHAM-15P1 5' end, mRNA sequence.
ACCESSION   BG584476
VERSION   BG584476.1 GI:13599540
KEYWORDS   EST.
SOURCE   Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM   Medicago truncatula/Glomus versiforme mixed EST library
            Eukaryota; mixed EST libraries.
REFERENCE   1 (bases 1 to 787)
AUTHORS   Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J.
            and Fraser,C.M.
TITLE   ESTs from roots of Medicago truncatula after colonization with

```
Db   454  GCCATGCAGGATGCTTCTGATGCTGTTAAAAGTGCCATGGTTAGTGTGTAGGATTGGAC 513
Qy   138  ValGluValIleGluGluAlaCys----------GlnLysAlaSerGluLeuGlyVal 153
          |||         ::::::|         |||
Db   514  TCAGAAAAGTCCAGCAGTGTGTGATGCACGGAAATCAGGAGGTGCTGAAGCTGAGAAG 573
Qy   154  ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyValValAla 173
          |||     |||                :::    |||    :::  ||| :||
Db   574  GTTCAGATTGCCAATTACCATGTCCAGGAAACTATGCTGTCTCTGGAGGCATAAAAGGA 633
Qy   174  ValAspArgAlaValAlaGluGluGlnGluAlaGlyAlaLysArgLeuIleProLeuLys 193
          ::::::|                       :::  |||        :::
Db   634  ATAGAACGCGGTGGAAGCCAAAGCAAAGTCTTTCAAGGCTCGAATGACTGTGCGCTTAGCT 693
Qy   194  ValSerGlyProPheHisThrAlaLeuLeuGluProAlaSerGlnLysLeu 210
          :::::::|
Db   694  GTTGCGGGAGCTTTCCATACTAGTTTTATGGAACCTGCCGTGTCAAGATTG 744
```

```
RESULT 2
BQ818656/c
LOCUS       BQ818656                   511 bp    mRNA    linear    EST 01-AUG-2002
DEFINITION  1030072D02.y1 C. reinhardtii CC-1690, Deflagellation (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BQ818656
VERSION     BQ818656.1    GI:22068897
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1030
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1..511
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Deflagellation
                     (normalized), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
                     XhoI; Deflagellation library, constructed by John Davies
                     and Jeffrey McDermott, combines cDNAs from CC-1690 cells
                     which had been re-synthesizing flagella for 15, 30 and 60
                     min after being deflagellated by pH shock. PolyA mRNA was
                     purified from each sample, pooled and cDNA synthesized.
                     The cDNA was directionally cloned into lambda Zap II
                     (Stratagene) in the EcoRI (5') and XhoRI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al., (1996) Genome Research 6: 791-806."
BASE COUNT      107 a    148 c    126 g    130 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:             1.93e-31
Score:                 324.50
Percent Similarity:    63.25%
Best Local Similarity: 41.57%
Query Match:           21.38%
DB:
```

```
Length:        511
Matches:       69
Conservative:  36
Mismatches:    60
Indels:        1
Gaps:          1
```

```
JOURNAL     Glomus versiforme, 2001
            Unpublished (2001)
COMMENT     Contact: Harrison M.J.
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73401
            Tel: 580-223-5810
            Fax: 580-221-7380
            Email: mjharrison@noble.org
            Noble EST name: N379946e TIGR sequence name: MTDBQ85TK More
            information is available at: http://www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1..787
                     /organism="Medicago truncatula/Glomus versiforme mixed EST
                     library"
                     /cultivar="Medicago truncatula genotype A17"
                     /db_xref="taxon:l19092"
                     /clone="pMHAM-15P1"
                     /clone_lib="MHAM"
                     /tissue_type="roots colonized with Glomus versiforme"
                     /dev_stage="roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The library was
                     made from a mixture of RNA from each of these stages."
                     /lab_host="E. coli strain XlOLR"
                     /note="Vector: pBluescript SK-; Site 1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     roots harvested at 10, 17, 22, 31 and 38 days
                     post-inoculation with Glomus versiforme. The cDNA was
                     directionally ligated into the Unizap XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-Zap phage using Ex-assist
                     helper phage and propagated in XlOLR cells."
BASE COUNT      204 a    154 c    206 g    223 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:             1.93e-31
Score:                 327.00
Percent Similarity:    55.30%
Best Local Similarity: 36.87%
Query Match:           21.54%
DB:
```

```
Length:        787
Matches:       80
Conservative:  40
Mismatches:    85
Indels:        12
Gaps:          5
```

```
US-09-308-397-2 (1-306) x BG584476 (1-787)

Qy     5  AlaPheLeuAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp 24
          |||||||||||  ||||  |||||::::::|||  :::||||:||
Db    97  GCATTTCTCTCCCCGGCAGGGTGCACAAGCTGTTGGAATGGGAAAAGAA--GCTCAA 153

Qy    25  GlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValGluGlyTyrAsp--- 43
          |||                :::|||    ::::||||  :||||:::::|::|
Db   154  AGTGTTCCTGCTGCTGCCACTTTGTACAAGGCCAATGAGATTTTGGGGTTTGAТCCT 213

Qy    44  LeuArgTyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGlnPro 63
          :::|||        ::::::||||        :::||||
Db   214  CTTGATGTGATGCATTAATGGGCCAAAGGAAAAACTGAATTCAACAGTTATTAGTTCAGCCT 273

Qy    64  AlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeu------GlnGluLysGlyTyr 81
          ||||:::                  |||:::|||        :::|||
Db   274  GCCCTGTATGTCACAAGTCTTGCTGCTGTTGAGCTACTTCGGGCACGTGAGGAGGGTGAG 333

Qy    82  Gln--------ProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeu 97
          :::        |||:::|
Db   334  CAGATTATTGATTCGTTGATGTTACAATGCCGTTTCAGCCTGGGAGAAATACTTCTCTG 393

Qy    98  ValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLysArgGlyAla 117
          |||         :::|||||||||  :||  :::      |||
Db   394  GCATTTGCTGGCGGCTTTCAGCTTGAAGATGGACTGAAAATTGGTGAAACTGAGGGTGAA 453

Qy   118  TyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrPro 137
          |||:::        ||| :::|||
```

```
BASE COUNT      118 a    161 c    126 g    133 t    6 others
ORIGIN

Alignment Scores:
  Pred. No.:                         Length:            544
  Score:              4.26e-30       Matches:            79
  Percent Similarity: 314.50         Conservative:       33
  Best Local Similarity: 61.54%      Mismatches:         67
  Query Match:         43.41%        Indels:              4
  DB:                  20.72%        Gaps:                1
                       9

US-09-308-397-2 (1-306) x AU069821 (1-544)

Qy   107 AspAlaValAlaLeuValAlaAlaLysArgGlyAlaTyrMetGluAlaAlaProAlaAsp 126
         |||||:: |||||  ||| :||| :  ||| | :::||:|| |   || ||||||| |||
Db   541 GATGGGTGCGTCGTTGGTTGCTGTCAGGTCGGTTCTTCCGGTAAACTTTAAC 482

Qy   127 SerGlyLysMetValAlaValLeuAsnThrProValGluVal-IleGluGluAlaCysGl 146
         ::: ||| :   ::|||| || :::::  |:: ||| ||| ||| ||||||||||||||||
Db   481 ACGGGCGCTATGNCGGCCAATCATCGGTTGGAATGATCGTCATTGCGAAAGCGTGTNA 422

Qy   146 nLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGlnIleVa 166
         ::: || :||||||| || |||  |  |||| ||| ||| || ||||||||||||:::::
Db   421 AGAAGCTGCGAAGGTCAGGTCGTTTCTTCCGGTAAACTTTAACTTTCCCGGA-CAGGTGGT 363

Qy   166 lIleAlaGlyValValAlaValAlaAspArgAlaValGluLeuLeuGlnGluAlaGlyAl 186
         |||||||||||||||| || || ||||| ||||||||| :||||||||:::
Db   362 TATTCCCGGTCATAAAGAAGCGGTTGAGCGTGCTGCCTGCTGTAAAGCGGGGGGC 303

Qy   186 a-LysArgLeuIleProLeuLysValSerGlyValProPheHisThrAlaLeuGluProA 206
         : |||  ||||| ::: ||||| |  ||:  |||| :: |||| |
Db   302 NAAAACGCGCGGTCGCCGTTACCAGTGACCGGTACCGTCCACTGTGCGCTGATGAAACCAG 243

Qy   206 laSerGlnLysLeuAlaGluThrLeuAlaGlnValSerPheSerAspPheThrCysProL 226
         |:: :: |||||||| ||| ||: ||| |  ||||| |   |||||||||||||||:::
Db   242 CAGCCGACAAACTGCCAGTAGAATTAGCGAAAATCACCTTAAGCACCAACAGTTCCTG 183

Qy   226 euValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp---IleAlaGlnLeuLeuT 245
         :: || :: |||:|: |||| || || || |:|||||| ||||| ||| |||
Db   182 TTGTGAATAACGTTGATGTGAAATGCGAAACCAATGGTGATGCCATCCGTGACGCACTGG 123

Qy   245 hrArgGlnValLysGluProValArgPheTyrGluSerIleGlyValMetGlnGluAlaG 265
         ||| ::: ::||||| |||| ||| |:||: ||::::||:::
Db   122 TACGTCAGTTGTATAACCCGGTTCAGTGACGAAGTCGTTGAGTACATGGCAGCGCAAG 63

Qy   265 lyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSerGlyPheValLysLysI 285
         ||::: |||||||  |||| ||||| ||||||||||||||||||||||||||||||||
Db    62 GCGTAGAACATCTCTATGAAGTCGGCCCCGGCAAAGTGCTTACTGGCCTGACGAAACGCA 3

Qy   285 le 285
         ||
Db     2 TT 1
```

RESULT 4

```
AJ500905

LOCUS       AJ500905               647 bp    mRNA            linear   EST 09-AUG-2002
DEFINITION  AJ500905 MTGIM Medicago truncatula cDNA clone mtgmaccl12002le04,
            mRNA sequence.
ACCESSION   AJ500905
VERSION     AJ500905.1  GI:22081838
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Krajinski,F., Manthey,K., Bartelsmeier,V., Meyer,F., Bartels,D.,
            Bekel,T., Linke,B., Franken,P., Kuester,H., Perlick,A.M. and
            Puehler,A.
  TITLE     Detection of transcript sequences from mycorrhizal roots of the
```

```
US-09-308-397-2 (1-306) x BQ818656 (1-511)

Qy   141 IleGluGluAlaCysGlnLysAlaSerGlnLeuGlyValValThrProAlaAsnTyrAsn 160
         |||  |||||||||||||||:::|| |||| ::|||||||||||||||||||||||||
Db   511 ATTGCGAAAGCCGTGAAGAAGCTGCCAGAAGGTCAGGTCGTTTCCGGTAAACTTTAAC 452

Qy   161 ThrProAlaGlnIleValIleAlaGlyValValAlaValAlaValAlaAspArgAlaGluLeu 180
         |||||||||:::::  |||||||||||||||||||||||| |||  |||
Db   451 TCTCCGGACAGGTGGTTATTGCCGGTCATAAAGAAGCGGTTGAGCGTGCTGGCCTGCC 392

Qy   181 LeuGlnGluAlaGlyAlaLysArgLeuIleAlaLeuLysSerGlyProPheHisThr 200
         :::  |||:||:|||||||||||||||||||||||||||||||||||||||||
Db   391 TGTAAAGCGGCGGGGGCGCAAAACGGCGCGGTCGCCGTTACCAGTGACGTACCGTCTCACTGT 332

Qy   201 AlaLeuGluProAlaSerGlnLysAlaGluThrLeuAlaGlnValSerPheSer 220
         ||||||||||| ||||||||||||||||||||| ||||||||||||||||:::
Db   331 GCGCTGATGAAACCAGCAGCCGACAAACTGCCAGTAGAATTAGCGAAAATCACCTTTAAC 272

Qy   221 AspPheCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp--- 239
         |||||| |  |  :: :::  :::
Db   271 GCACCAACAGTTCCTGTTGTGAATAACGTTGATGTGAAATGCGAAACCAATGGTGATGCC 212

Qy   240 IleAlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGly 259
         |||:::  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   211 ATCCGTGACGCACTGGTACGTCAGTTGTATAACCCGGTTCAGTGGACGAAGTCGTTGAG 152

Qy   260 ValMetGlnGluAlaGlyIleSerAsnPheIleGluIleGlyProGlyLysValLeuSer 279
         |||:::|||:::||:|||:::|||||| |||  :: ||||||||||||||||||:::
Db   151 TACATGCAGCGCAAGGCGTAGAACATCTCTATGAAGTCGGCCCCGGCAAAGTGCTTACT 92

Qy   280 GlyPheValLysLysIleAspPheIleThrAlaHisLeuAlaHisValGluAlaGlnAlaSer 299
         |||  ||||||||||||  ||| ||| :: :::  ::: :::
Db    91 GGCCTGACGAAACGCAATGTCGACACCGTGAGACCGCCGCTCGGCGTGAAGCGAACCTTCAGCG 32

Qy   300 LeuValAlaLeuGluGlu 305
         ::: ||| ||| |||||
Db    31 ATGGCAGCGGCGGCTCGAG 14
```

RESULT 3

```
AU069821/c

LOCUS       AU069821               544 bp    mRNA            linear   EST 02-APR-2002
DEFINITION  AU069821 Rice panicle at flowering stage Oryza sativa (japonica
            cultivar-group) cDNA clone E3726_7A, mRNA sequence.
ACCESSION   AU069821
VERSION     AU069821.1  GI:5004646
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Sasaki,T. and Yamamoto.K.
  TITLE     Rice cDNA from panicle at flowering stage
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1..544
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="E3726_7A"
                     /clone_lib="RIce panicle at flowering stage"
                     /dev_stage="flowering stage"
                     /note="Organ: panicle; Rice cDNA from panicle at flowering
```

model mycorrhiza Medicago truncatula genotype A17 - Glomus
intraradices using the approach of an EST genome project based on
an SSH library

JOURNAL   Unpublished (2002)
COMMENT   Contact: Krajinski F
           LG Molekulargenetik
           Herrenhaeuser Str. 2 D-30419 Hannover, Germany.

FEATURES   Location/Qualifiers
     source   1..647
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="mtgmacc120021e04"
                     /clone_lib="MTGIM"
                     /tissue_type="mycorrhizal roots"
                     /dev_stage="3 weeks after inoculation"
                     /note="Vector: pGEM-Teasy; genotype A17; cDNA was prepared
                     from total RNA using the SMART PCR cDNA system (Clontech)
                     from roots harvested three weeks after inoculation with
                     Glomus intraradics. This cDNA was used as tester in a
                     Suppression Subtractive Hybridization (SSH). The SSH-cDNA
                     fragments were generated using the SSH-adaptor sequences
                     ctaatacgactcactatagggctcgagcggccgcccgggcagt and
                     ctaatacgactcactatagggcagtggtccgcggccggt (Clontech) and
                     ligated after Suppression Subtractive Hybridization in to
                     the pGEM-Teasy vector from Promega. Plasmids containing
                     cDNA inserts were propagated in E. coli TOP 10F' cells
                     (Invitrogen)."

BASE COUNT   174 a   117 c   181 g   175 t
ORIGIN

Alignment Scores:
Pred. No.:   1.24e-27   Length: 647
Score: 296.50   Matches: 72
Percent Similarity: 53.59%   Conservative: 40
Best Local Similarity: 34.45%   Mismatches: 86
Query Match: 19.53%   Indels: 11
DB: Gaps: 4

US-09-308-397-2 (1-306) x AJ500905 (1-647)

Qy   35 ArgAlaSerGlnValLeuGlyTyrAsp--LeuArgTyrLeuIleAspThrGluAsp 53
           ||:||||:||||:::|||||||:||| :::||||||||||||||||||||
Db   9 AAAGCCAATGAGAGATTTGGGGTTTGATCTTCTTGATGTATGCATTAATGGCCAAGGAA 68

Qy   54 LysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyr 73
           ||||||||||:||||||||||||||:||||||:|||||||:|||||||||:
Db   69 AAACTGAATTCAACAGTTATTAGTCAGCCCTGCCCTGTATGTCACAAGTCTTGCTGCTGTT 128

Qy   74 ArgLeuLeu-----GlnGluLysGlyTyrGln-------ProAspMetValAla 87
           |||||| :::||||||||| :||:|
Db   129 GAGCTACTTCGGGCACGTGAGGAGGTGAGCAGATTATTGATTGTTGATGTTACATGC 188

Qy   88 GlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyValAlaLeuAspPheGluAsp 107
           :||||||||:|||:||||||||:::::||||:||||||
Db   189 GGTTTGAGCCCTGGAGAATATACTGCTTCCTGGGGCTTTCAGCTTTGAAGAT 248

Qy   108 AlaValAlaLeuValAlaLysArgGlyAlaTyrMetGluAlaAlaProAlaAspSer 127
           :::||||||||||||||:|||||||| |||:|||||||:
Db   249 GGACTGAAATTGGTGAACACTGAGGGGTGAAGCCATGCAGGATGCTTCTGTGTCTGTTAAA 308

Qy   128 GlyLysMetValAlaValAsnThrProValGluValIleGluGluValAlaCys------ 145
           |||||||:||| ::|||:|||
Db   309 AGTGGTATGGTTAGTGTGGATGGGGAGGATGGACTGGACTCCAGCAGTTGTGTGATGCA 368

Qy   146 ------GlnLysAlaSerGluLeuGlyValValThrProAlaAsnThrArgProAla 163
           :::|||||| ::||||||:::|||
Db   369 GCGAATCAGGAAGTGCCTGAAGCTGAGAAGGTTCAGATTGCCAATTACCTATGTCCAGGA 428

Qy   164 GlnIleLeuAlaGlyGluValAlaValAlaValAspArgAlaValGluLeuLeuGlnGlu 183
           :::::|||||| ::: :::
Db   429 AACTATGCTGTCTCTGGAGGCATAAAAGGATAACGCGGTGAAGCCAAAGTCT 488

Qy   184 AlaGlyAlaLysArgLeuIleProLeuLysLeuValSerGlyProPheHisThrAlaLeuLeu 203

Db   489 TTCAAGGCTGGAATGACTGTGCCTTAGCTGTTGCGGGAGCTTTCATACTAGTTTTATG 548
           |||:||| ::: ||| ||||||||:|||
Qy   204 GluProAlaSerGlnLysGlyLeuAlaGluThrGluGlnValSerPheSerAspPheThr 223
           :::||| :||||
Db   549 GAACCTGCGGTGCAAGATTGGAAGCTGCATCATGGCATCGACAGACATTAGAACCCAAGA 608

Qy   224 CysProLeuValGlyValGlyAsnThrGluAla 232
           :||
Db   609 ATACCAGTCATTTCCAATGTGAATGCT 635

RESULT 5
AJ498681
LOCUS   AJ498681   690 bp   mRNA   linear   EST 09-AUG-2002
DEFINITION   AJ498681 MTPOSE Medicago truncatula cDNA clone mt--acc95207e02,
             mRNA sequence.
ACCESSION   AJ498681
VERSION   AJ498681.1   GI:22089124
KEYWORDS   EST.
SOURCE   barrel medic.
  ORGANISM   Medicago truncatula
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
             Medicago.
REFERENCE   1 (bases 1 to 690)
  AUTHORS   Firnhaber,C., Bartelsmeier,V., Meyer,F., Bartels,D., Bekel,T.,
             Linke,B., Puehler,A. and Kuester,H.
  TITLE   Determination of transcript sequences from developing pods
             including seeds of Medicago truncatula genotype A17
  JOURNAL   Unpublished (2002)
COMMENT   Contact: Kuester H
             Lehrstuhl fuer Genetik
             Universitaet Bielefeld
             Postfach 100131, D-33501 Bielefeld, Germany.
FEATURES   Location/Qualifiers
     source   1..690
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="mt--acc95207e02"
                     /clone_lib="MTPOSE"
                     /tissue_type="pods including seeds"
                     /dev_stage="different stages of development"
                     /note="Vector: pGEM-T; Site_1: PstI; Site_2: SphI;
                     genotype A17; cDNA was prepared from polyA+ enriched RNA
                     from developing pods including seeds harvested at
                     different stages of development. The cDNA was
                     directionally ligated by MediGenomix into the pGEM-T
                     vector from Promega using GCATCGCCAGGCGGCGCCGACATG and
                     CTGCAGGCCATATGGCCGGG adapters. Plasmids containing cDNA
                     inserts were propagated in E. coli DH10B cells."

BASE COUNT   188 a   135 c   190 g   177 t
ORIGIN

Alignment Scores:
Pred. No.:   3.4e-27   Length: 690
Score: 293.50   Matches: 73
Percent Similarity: 50.88%   Conservative: 42
Best Local Similarity: 32.30%   Mismatches: 100
Query Match: 19.33%   Indels: 11
DB: Gaps: 4

US-09-308-397-2 (1-306) x AJ498681 (1-690)

Qy   52 GluAspLysLeuAsnGlnThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAla 71
           ::|||::|||||:|||::|||||||||:||||:||||||||:|||||||||:
Db   11 AAGGAAAAACTGAATTCAACAGTTATTAGTCAGCCTGCCCTGTATGTCACAAGTCTTGCT 70

Qy   72 IleTyrArgLeuLeu------GlnGluLysGlyTyrGln--------ProAspMet 85
           |||||||||| |||||||:|| ::
Db   71 GCTGTTGAGCTACTTCGGGCACGTGAGGAGGTGAGCAGATTATTGATTCTTGTTGATGTT 130

Qy   86 ValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSerGlyValAlaLeuAspPhe 105

```
                                                                    XhoI; This library, constructed by John Davies and Jeffrey
                                                                    McDermott, combines cDNAs from CC-1690 cells grown to
                                                                    mid-log phase in TAP (acetate-containing) medium in the
                                                                    light, TAP medium in the dark, HS (minimal) medium in
                                                                    ambient levels of CO2 and HS medium bubbled with 5% CO2.
                                                                    PolyA mRNA was purified from each sample, pooled and cDNA
                                                                    synthesized. The cDNA was directionally cloned into lambda
                                                                    ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                                                                    pBluescript II SK- plasmids were excised from the lambda
                                                                    ZAP clones by superinfection with Exassist (Stratagene)
                                                       Qy.  146  ----------GlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThr 161
                                                                    phage. The library was normalized using method 4 described
                                                                    in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT    128 a      177 c     214 g     128 t
ORIGIN

Alignment Scores:
Pred. No.:              6.42e-27                 Length:            647
Score:                  291.00                   Matches:           69
Percent Similarity:     58.72%                   Conservative:      32
Best Local Similarity:  40.12%                   Mismatches:        61
Query Match:            19.17%                   Indels:            10
DB:                     12                       Gaps:              5

US-09-308-397-2 (1-306)  x  BG851744 (1-647)

Qy    7  LeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAspGlnTyr  26
            ||||||||::|::||   :||||||||||||||||||||||  :::||||||||
Db   11  TTGTTCCCGGCCCAGGGTGCGCAGAGCGTGGCATGGCCAAGGACCTGTGGCGGCGACCGTG  70

Qy   27  ProIleValLysGluThrIleAspArgAlaSerGlnLysValLeuGlyTyrAsp---LeuArg  45
            :||||||||||||:||    |||||||:::||||::||||||||||||  :::||||  :
Db   71  CCGAAGGCCAAGGACGATGTTCGACAAGGCGTCTGAGATCCTGAGACGTCTACGACCTGCTGAAG 130

Qy   46  TyrLeuIleAspThrGluGluAspLysLeuAsnGlnThrArgThrGlnProAlaIle  65
            ||||  ||||||||||  :::||||||||||||||||||||||||
Db  131  GTGTGCGTGGAGGGCCCCAAGGAGGAAGCTGGACGACCACGCCGTGAGTCAGCCGCTATC 190

Qy   66  --LeuAlaThrSerValAlaIleTyrArgLeuLeuGlnLeuGlyLysGlyTyrGlnPro--  83
            :::|||||||  ||  ::||||||
Db  191  TACGGTGGCCCTCGCTGGCCGCGGAGAGCTGGCGGCGACGAGGGCCAGGCCGCCCATT 250

Qy   84  ------------AspMetValAlaGlyLeuSerLeuGlyGluTyrSerAlaLeuValAlaSer 100
            :::   |||||  ||||||
Db  251  GACGCCATTGACGTGGCCGTGCGGCCTGCTGCTGGGCGAGTACAGCGCCCTGGCCTTCGCT 310

Qy  101  GlyAlaLeuAspPheGluAspArgAlaLysArgGlyAlaTyrMetGlu 120
            |||||||:::  ::|||||||||:  :  ||||||
Db  311  GGTGCCATGAGCTTCGAGGACGCGCTCCGGCGCTGGGTGAAGCTCCATGCAG 370

Qy  121  GluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProValGluVal 140
            :::
Db  371  GCGGGCTGCTGACGCGCCAGCCCAGCATGGTTCCGTCATTGGGCTGGACAGCGCCAAG 430

Qy  141  IleGluGluAlaCysGlnLysAlaSerGlu---LeuGlyValValThrPro------- 156
            ||  :::|||||||||:||:||  ||||||||
Db  431  GTGGCGGAGGCTGCCAAGGTGCCAAGGTTCAGGAAAACGCCCGGTCAGATC 490

Qy  157  AlaAsnTyrAsnThrProAlaGlnIleValIleAla 168
            :::||||
Db  491  GCCAACTACCTGCTGCCCCTGTCACTACGGCGTGTCG 526
```

```
Db  131  ACATGCGCGTTTGAGCCTGGAGAGATATACTGCTCTGCGCATTTGCTGGGGCTTTCAGCTTT 190
            :::||||||
Qy  106  GluAspAlaProAlaValAlaLeuValAlaLeuAlaValAlaArgGlyAlaTyrMetGlyLeuAlaAlaProAla 125
            ||||:::
Db  191  GAAGATGGACTGAAATTGGTGAAACTGAGGGTGAAGCCATGCAGGATGTGTTCTGATGCT 250

Qy  126  AspSerGlyLysMetValAlaValLeuAsnThrProValGluValIleGluGluAlaCys 145
            ||||||:::::||::|:
Db  251  GTTAAAAGTGCCATGGTTAGTGTGGTAGGAATTGGACTCAGAAAAAGTCCAGACCAGTTGT 310

Qy. 146  ----------GlnLysAlaSerGluLeuGlyValValThrProAlaAsnTyrAsnThr 161
            ||:::|||||
Db  311  GATGCGAGCGAATCAGGAAGTGCCTGAAGCTGAGAAGGTTCAGATTGCCAATTACCTATGT 370

Qy  162  ProAlaGlnIleValIleAlaGlyAlaGluValAlaLeuAlaValAspArgAlaValGluLeuLeu 181
            ||||      :::|||||||
Db  371  CCAGGAAACTATGCTGTCTCTGGAGGCATAAAAGGATAGAGAGCCGGTGGAAGCCAAAGCA 430

Qy  182  GlnGluAlaGlyAlaLysArgGluIleProLeuLysProLeuSerGlyValProPheHisThrAla 201
            :::          |||::  :|:||||:::|||
Db  431  AAGTCTTTCAAGGCTCCAATGACTGCGCCTAGCTGTTGCGGGAGCTTTCCATACTAGT 490

Qy  202  LeuLeuGluProAlaSerGlnLysGlnLysLeuAlaGlyLeuAlaGluThrLeuAlaGlnValSerPheSerAsp 221
            |||
Db  491  TTTATGGAACCTGCCGTGTCAAGATTGGAAGCTGCATTGGCATCGACAGACATTAGAACC 550

Qy  222  PheThrCysProLeuValGlyAsnThrGlyAlaAlaValMetGlnLysGluAsp---Ile 240
            |||||||||
Db  551  CCAAGAATACCAGTCATTTCCAATGTGAATTCTCTGCCGCACACAGATCCTGACACAATA 610

Qy  241  AlaGlnLeuLeuThrArgGlnValLysGluProValArgPheTyrGluSerIleGlyVal 260
            :::
Db  611  AAGAGGATATCCGCACGACAGTTACTTCCCGTTCAATGGAAACAACTGTGAAGACT 670

Qy  261  MetGlnGluAlaGlyIle 266
            :::|||
Db  671  CTTCTATCCAAGGGGTG 688
```

```
RESULT 6
BG851744
LOCUS       BG851744                 647 bp    mRNA    linear    EST 29-MAY-2001
DEFINITION  1024032B11.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG851744
VERSION     BG851744.1  GI:14223928
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu
FEATURES             Location/Qualifiers
     source          1..647
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
```

```
RESULT 7
BE404963
LOCUS       BE404963                 585 bp    mRNA    linear    EST 21-JUL-2000
DEFINITION  WHE1207_B01_D012S Wheat etiolated seedling root cDNA library
            Triticum aestivum cDNA clone WHE1207_B01_D01, mRNA sequence.
ACCESSION   BE404963
VERSION     BE404963.1  GI:9364431
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
```

```
Qy     97 LeuValAlaSerGlyAlaLeuAspPheGluAspAlaAlaValAlaAlaLeuValAlaLysArgGly 116
             ||||| :::   ||||||||
Db    355 CTTGCATTGTGGTGCCTTAGCTTGAGGATGGTCTGAAGCTTGTCAAGCTTAGAGGA 414

Qy    117 AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
             |||  :::||| ||:::  ||||||::||
Db    415 GAAGCCATGCAGGATGCCTCAGACCGCTGCCAATAGTGCGATGGTTAGTGTAATTGGTCTA 474

Qy    137 ProValGluValIleGluGluAlaCysGlnLysAlaSer------------GluLeuGly 152
             ||||| :::||: ::|:|:
Db    475 GATTCAGAAAAGGTGCAACAATTATCGATGCTGCAAATGAGGATGTGGATGAAAAGGAA 534

Qy    153 ValValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGly 169
             :::::  |||
Db    535 AGAGTTCAAATAGCAAATTTCTCTGTCCTGTAACTATGCAGTTTCTGGT 585

RESULT 8
BF512296
LOCUS           BF512296                    491 bp    mRNA     linear    EST 07-DEC-2000
DEFINITION      UI-H-BW1-amb-e-04-0-UI.s1 NCI_CGAP_Sub7 Homo sapiens cDNA clone
                IMAGE:3069391 3', mRNA sequence.
ACCESSION       BF512296
VERSION         BF512296.1  GI:1159475
KEYWORDS        EST.
SOURCE          human.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 491)
  AUTHORS       NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
                Tumor Gene Index
  JOURNAL       Unpublished (1997)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Oligo-dT track not found, Not I site shown in beginning of sequence
                is likely internal to the message. cDNA Library Preparation: M.B.
                Soares Lab Clone distribution: NCI-CGAP clone distribution
                information can be found through the I.M.A.G.E. Consortium/LLNL at:
                www-bio.llnl.gov/bbrp/image/image.html
                Seq primer: M13 Forward
                POLYA=No.
FEATURES            Location/Qualifiers
     source         1..491
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:3069391"
                    /clone_lib="NCI_CGAP_Sub7"
                    /lab_host="DH10B (life Technologies)"
                    /note="Vector: pTT3D-Pac (Pharmacia) with a modified
                    polylinker; Site 1: Not I; Site 2: Eco RI; NCI_CGAP_Sub7
                    NCI_CGAP_Sub7 library had 12 million recombinants. A
                    single-stranded DNA preparation of NCI_CGAP_Sub6 was used
                    as a tracer in a subtractive hybridization with a driver
                    comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1  LLAM
                    3334-3337, 3682-3683, 3798-3803-(IMAGE CloneIDs
                    1322376-1323911, 1456008-1456775,1500552-1502855);
                    NCI_CGAP_Kids pool 1 LLAM 3338-3342,3722-3725, 3776-3778
                    (IMAGE CloneIDs 1323912-1325831, 1471168-1472903,
                    1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
                    3851-3854 (IMAGE CloneIDs 1414920-1417991, 1520904-1522439
                    ); NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720,
                    3733-3735  (7IMAGE CloneIDs 1257096-1258631,1469064-1470983
                    ,1475592-1476743); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459,
                    2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
                    1101192-1101959, 1217928-1220615); NCI_CGAP_Co10 pool 1
                    LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255
                    ,1144584-1145351). (6% of the driver population), plus a
                    pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE
                    CloneIDs 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE
                    CloneIDs 2710536-2712455). (4% of the driver population
                    ), plus a pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE
```

```
REFERENCE     ; Triticeae; Triticum.
  AUTHORS       Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
                ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
                Seaton,C.L. and Tong,J.C.
  TITLE         The structure and function of the expressed portion of the wheat
                genomes
  JOURNAL       Unpublished (2000)
COMMENT         Contact: Olin Anderson
                US Department of Agriculture, Agriculture Research Service, Pacific
                West Area, Western Regional Research Center
                800 Buchanan Street, Albany, CA 94710, USA
                Tel: 5105595773
                Fax: 5105595818
                Email: oanderson@pw.usda.gov
                Sequence have been trimmed to remove vector sequence and low
                quality sequence with phred score less than 20
                Seq primer: Stratagene SK primer.
FEATURES            Location/Qualifiers
     source         1..585
                    /organism="Triticum aestivum"
                    /cultivar="Chinese Spring"
                    /db_xref="taxon:4565"
                    /clone="WHE1207_B01_D01"
                    /tissue_type="Root"
                    /dev_stage="Five day old etiolated seedling"
                    /lab_host="E. coli SOLR"
                    /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                    Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized
                    , germinated and grown aseptically in the dark at room
                    temperature on filter paper with water, nystatin and
                    cefotaxime in covered crystallization dishes. Roots were
                    harvested. The tissue, total RNA, and poly(A) RNA were
                    prepared, a cDNA library was made, and the cDNA clones
                    were in vivo excised to give pBluescript phagemids in the
                    TJ Close lab (Choi, Close, Fenton) at the University of
                    California, Riverside. Plasmid DNA preparations and DNA
                    sequencing were performed in the OD Anderson lab (all
                    other authors)."

BASE COUNT      154 a     121 c     157 g     153 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                  7.27e-27          Length:           585
Percent Similarity:     290.00            Matches:          72
Best Local Similarity:  58.19%            Conservative:     31
Query Match:            40.68%            Mismatches:       62
DB:                     19.10%            Indels:           12
                        10                Gaps:             5

US-09-308-397-2 (1-306) x BE404963 (1-585)

Qy      4 ThrAlaPheLeuPheAlaAlaGlyGlnAlaGlnTyrLeuGlyMetGlyArgAspPheTyr 23
            ||||||||||:||||:|| :::  ::|:||:::
Db     58 ACCGCATTCCTTTTCCCCGGCCAGGTGCTCAGGCTGTTGGAATGGGTAAAGAGGGCTCTT 117

Qy     24 AspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnIleValLeuGlyTyrAsp 43
            :::       ||||:||  ||||:::
Db    118 AAT--GTTCGAGCAGCTGCAGAACTATTTGATAAGGCAAATGATATACTTGGCTATGAC 174

Qy     44 LeuArgTyrLeu--IleAspThrGluGluAspLysLeuAsnGlnThrArgTyrThrGln 62
            ||||||    |||||:| |||:::
Db    175 TTGTTGAATCTTTGCCATCGATGGACCAAAAGAAAAGCTGAATTCAACAGTGATCAGTCAG 234

Qy     63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeu------GlnGluLysGly 80
            ||||||||||  ||  |:|| ::|:
Db    235 CCAGCTATATATNTACCAGCCTTGCAGCTGTAGAAGTGTTGCGCGACGTGAAGAGAGGC 294

Qy     81 Tyr---------GlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAla 96
            |||::|
Db    295 CAATCTGTAATTAACTCCGTAGATGTCACATGTGGTCTCAGCTTGGGAGAATTACCGCG 354
```

REFERENCE
AUTHORS   CloneIDs 2712456-2723591) (10% of the driver population),
    plus a pool of 5,472 clones from NCI_CGAP_Sub4 (IMAGE
    CloneIDs 2723592-2729326) (40% of the driver population),
    plus a pool of 4032 clones from NCI_CGAP_Sub6 (IMAGE
    CloneIDs 2728969-2733190) (40% of the driver population).
    Subtraction was performed as previously described (Bonaldo
    , Lennon & Soares (1996): Normalization and Subtraction:
    Two Approaches To Facilitate Gene Discovery. Genome
    Research 6, 791-806.
    TAG_SEQ=None found"

BASE COUNT   90 a   152 c   167 g   82 t
ORIGIN

Alignment Scores:
Pred. No.:   6.23e-27   Length: 491
Score: 289.50   Matches: 65
Percent Similarity: 56.17%   Conservative: 26
Best Local Similarity: 40.12%   Mismatches: 64
Query Match: 19.07%   Indels: 7
DB: Gaps: 2

US-09-308-397-2 (1-306) x BF512296 (1-491)

Qy   117 AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
Db   3 GCCGCCATGCAGGAAGCCGTTCGGTCGGTACCGGCGCCCATCCTCCGCATG 62

Qy   137 ProValGluValIleGluGluAlaCysGlnLys------------AlaSerGlu 150
Db   63 GACGCCCAGCGCGTGATTGCTCGCAGAGGCCAAGCCACATTGGCGCCGGCACT 122

Qy   151 LeuGlyValValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGlyGlu 170
Db   123 TTGGAGGTGGTGGAAGCGGTGAACTTCAACGACCGGTGATCGCGGGGGAGC 182

Qy   171 ValAlaValAlaAspArgAlaValGluGluLeuGlnGlnAlaGlyAlaLysArgLeuIle 190
Db   183 AAGGCCGGCGGTCGACAAGGCCTGTGAGGTGCTCAAGGCCAACGGTGCGAAGCGCGCCTTG 242

Qy   191 ProLeuLysValSerGlyProPheHisThrAlaLeuLeuGluGluProAlaSerGlnLysLeu 210
Db   243 CCATTGCCGGTGTCGGCACCGTTCCATTCCAGCCTCATGAAGCCGCTGCCGAGCGTCTG 302

Qy   211 AlaGluThrLeuAlaGlnValSerPheSerAspPheThrCysProLeuValGlyAsnThr 230
Db   303 AAAGAAAAGCTGCCAATACGGTGTTTTCCCAGGCGCAGATCCCGGTGGTCAACACATC 362

Qy   231 GluAlaAlaValMetGlnLysGluAsp---IleAlaGlnLeuLeuThrArgGlnValLys 249
Db   363 GATGTGCGGTGGAAACCGATGCCAACCCGATGCGCATCCGCATGCTGTACCGCCAGGCATTC 422

Qy   250 GluProValArgPheTyrGluSerIleGlyValMetGlnValAlaGlyIleSerAsnPhe 269
Db   423 GGCCCCGGTGCGCTGGGTGGAATGTGTGCAGGCCATCAAGGGCGCTCGGCGGTCGACACGATT 482

Qy   270 IleGlu 271
Db   483 GTCGAA 488

RESULT 9
BE440632
LOCUS   BE440632   648 bp   mRNA   linear   EST 04-DEC-2001
DEFINITION   sp50b03.y1 Gm-c1043 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
          Gm-c1043-1638 5' similar to TR:O22916 O22916 PUTATIVE
          MALONYL-COA:ACYL CARRIER PROTEIN TRANSACYLASE ;, mRNA sequence.
ACCESSION   BE440632
VERSION   BE440632.1   GI:9440121
KEYWORDS   EST.
SOURCE   soybean.
ORGANISM   Glycine max
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

REFERENCE
AUTHORS   Glycine.
    1 (bases 1 to 648)
    Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
    ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
    Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
    ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
    ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
    ,R., Waterston,R. and Wilson,R.
    Public Soybean EST Project
TITLE   Public Soybean EST Project
JOURNAL   Unpublished (1999)
COMMENT   Contact: Shoemaker R/Public Soybean EST Project
    Public Soybean EST Project
    Washington University School of Medicine
    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: est@watson.wustl.edu
    This clone is available through: ResGen, Invitrogen Corp. 2130
    South Memorial Parkway Huntsville, AL 35801 For further information
    call: (800)-533-4363 or contact via email: ccu@resgen.com
    High quality sequence stop: 484.

FEATURES   Location/Qualifiers
    source   1..648
                /organism="Glycine max"
                /db_xref="taxon:3847"
                /clone="GENOME SYSTEMS CLONE ID: Gm-c1043-1638"
                /clone_lib="Gm-c1043"
                /tissue_type="Hypocotyl and Plumule, germinating seeds"
                /lab_host="DH10B"
                /note="vector: pT7T3Pac (Pharmacia); Site 1: EcoRI;
                Site 2: NotI; This cDNA library was constructed from mRNA
                isolated from hypocotyl and plumule tissues of seeds
                germinated for three days of the cultivar Williams.
                Complementary DNA was synthesized from mRNA using a primer
                consisting of a poly(dT) sequence with a NotI restriction
                site. EcoRI adapters were ligated to the blunt-ended cDNA
                fragments followed by digestion with EcoRI and NotI. The
                cDNA fragments were directionally cloned into the
                EcoRI-NotI restriction site of the pT7T3-Pac vector. The
                ligated cDNA fragments were transformed into DH10B host
                cells (Gibco BRL). This library was constructed by Dr.
                Randy Shoemaker."

BASE COUNT   189 a   133 c   172 g   151 t   3 others
ORIGIN

Alignment Scores:
Pred. No.:   2.69e-25   Length: 648
Score: 278.50   Matches: 67
Percent Similarity: 50.70%   Conservative: 41
Best Local Similarity: 31.46%   Mismatches: 92
Query Match: 18.35%   Indels: 13
DB: Gaps: 4

US-09-308-397-2 (1-306) x BE440632 (1-648)

Qy   91 LeuGlyGluTyrSerAlaLeuValAlaSerGlyAlaLeuAspPheGlnAspAlaAlaValAla 110
Db   16 CTGGGAGAATATACTGCTCTCGGCATTGCTGGGGCTTTGAAGATGGACTCAAG 75

Qy   111 LeuValAlaLysArgGlyAlaTyrMetGluGluValAlaAlaProAlaAspSerGlyLysMet 130
Db   76 TTGGTCAAACTGAGGGGTGAGGCCATGGAAGAAGTAGCGGCGGATGCCGCTGATGCTGCTAAAAGTGCTATG 135

Qy   131 ValAlaValLeuAsnThrProValGluValIleGluGluAlaCys---------Gln 146
Db   136 GTTAGTGTGATAGGACTGGACTCAGAGAAGGTCCAACAATTGTGTGATGCAGCCAATCAG 195

Qy   147 LysAlaSerGluGluLeuGlyValValThrProAlaAsnTyrAsnThrProAlaGlnIleVal 166
Db   196 GAAGTTCCTGAAGCTGAAGGTTCAGATTGCCAATTACCTGTCGCCAGGAAACTATGCT 255

Qy   167 IleAlaGlyGluValAlaValAlaAspArgAlaValGluGluLeuGlnGlnAlaGly--- 185
Db

```
Qy    4  ThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyr   23
            |||||||  |||| |:| |||||:::|  ||||| ||  ||| |||::|||:||
Db   34  ACCGCGTTCCTTTTCCCGGCCAGGGTGCTGGAATCCAAGGCAAGTGGAATGGTAAAGAAGACTCTT   93

Qy   24  AspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp   43
            :::  ::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db   94  AAT---GTTCGAGCAGCTGCAGAACTATTTGATAAGGCAAATGATATACTTGGCTACGAC  150

Qy   44  LeuArgTyrLeu---IleAspThrGluGluAsnGlnThrArgTyrThrGln   62
            |||  |||  ||||||||  |||| |||||||||||||||:|||:|||
Db  151  TTGCTGAATCTTTGCATCGATGGACCAAAGAAAAGTGAACTCAACAGTGATCAGTCAG  210

Qy   63  ProAlaIleLeuAlaThrSerValAlaIleTyrArgGluLeu------GlnGluLysGly   80
            |||||||  |||  ||||||  ||:|||||:|||  ::::||  |||||
Db  211  CCAGCTATATATGTTACCAGCCTTGCAGTCTGAGAGTGTTACGTGCACGGGAAGACGGT  270

Qy   81  Tyr----------GlnProAspMetValAlaGlyLeuSerLeuGlyGluTyrSerAla   96
                       ||:::  |||  ||||||||||  ||||||||||
Db  271  CAATCTGTAATTAACTCGTAGAGTGCACATGTGGTCTCAGCTTGGGAGAATATACCGCG  330

Qy   97  LeuValAlaSerGlyAlaLeuAlaAspPheGluAspProAlaValAlaLeuValAlaLysArgGly  116
            |||  |||  |||  |||||||  |::  |||
Db  331  CTTGCATTTGCTGCTGGTGCCTTTAGCTTTGGAGGATGGTCTGAAGCTTGCAAGCTTAGAGGA  390

Qy  117  AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr  136
            ::::|||  :::  |||:||||:||||||||||||  ::||||:|||:|||
Db  391  GAAGCCATGCAAGATGCCTCAGATGCGTGCCAATAGTGCGATGGCTAGTGTGATTGGTCTA  450

Qy  137  ProValGluValIleGluGluAlaCysGlnLysAlaSerGlu  150
            ::::||||||:::  ::::  ::||||||||
Db  451  GATTCAGAAAAGGTGCAACAATTATGCCATGCTGCAAATGAG  492
```

RESULT 11

```
BG444706
LOCUS        BG444706                850 bp    mRNA      linear   EST 15-MAR-2001
DEFINITION   GA__Ea0025E19f Gossypium arboreum 7-10 dpa fiber library Gossypium
             arboreum cDNA clone GA__Ea0025E19f, mRNA sequence.
ACCESSION    BG444706
VERSION      BG444706.1  GI:13354358
KEYWORDS     EST.
SOURCE       Gossypium arboreum
ORGANISM     Gossypium arboreum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
             Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE    1  (bases 1 to 850)
AUTHORS      Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
             ,D., Wood,T.C., Yu,Y., Leslie,K.A. and Wilkins,T.A.
TITLE        An integrated analysis of the genetics, development, and evolution
             of the cotton fiber
JOURNAL      Unpublished (2000)
COMMENT      Contact: Wing RA
             Clemson University Genomics Institute
             Clemson University
             100 Jordan Hall, Clemson, SC 29634, USA
             Tel: 864 656 7288
             Fax: 864 656 4293
             Email: rwing@clemson.edu
             Seq primer: TAATACGACTCACTATAGGG
             High quality sequence start: 3
             High quality sequence stop: 791
FEATURES             Location/Qualifiers
     source          1..850
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0025E19f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
```

```
Db  256  GTCTCCGGGGATTA-----------AAAGGAGTAGAAGTGCTGGAATCCAAGGCAAAG  303
Qy  186  -------AlaLysArgLeuIleProLeuLysValSerGlyValProPheHisThrAlaLeu  202
            ::|:::  |||||  ||::::::||:||  ||:||||
Db  304  TCTTTCAAGGCTCGAATGACGGTGCGCGCTAGCGTGTTGCTGGTGCTTTCCACACTAGTTTT  363
Qy  203  LeuGluProAlaSerGlnLysGlnAlaGlyLeuAlaGluThrLeuAlaGlnValSerPheSerAspPhe  222
            ::::|||||||  |||||  :::||    |||||
Db  364  ATGAAACCAGCTGCTGTCTGTCAAGGTTGGAGCAGCATTGGCAACAACAGAAATCAGAACCCCA  423
Qy  223  ThrCysProLeuValGlyAsnThrGluAlaAlaValMetGlnLysGluAsp---IleAla  241
            |||  |::||  ||||  |||    ||||
Db  424  AGATACCAGTCATCTCCAATGTGAAATGCACAGCCACATACAGATCCTGACTCAGTCAG  483
Qy  242  GlnLeuLeuThrArgGlnValLysGluThrIleGlyProValArgPheTyrGluSerIleGlyValMet  261
            ::::|||    |||  ::  |||  :::  :::||
Db  484  AAGATATTGGCATGCCANGTTACCTCACCTGTTCAATGGGAAACAACAGTGCAGACTCTT  543
Qy  262  GlnGluAlaGlyIleSerAsnPheIleGlyIleIleGlyProGlyLysValLeuSerGlyPhe  281
            |||||  :::||    ||||||||||    ||||
Db  544  CTAACCAAGGGTCTGAAGAAAAGTATGAACTCCGACCCCGGAAGGTTATTGCTGGAATC  603
Qy  282  ValLysLysIleAspGlnThrAlaHisThrAlaHisLeuAlaHisVal  294
            ||||||  :::  ||||
Db  604  ATCANAAGACGGACCAAGGTGCTGATATTGAANANATATT  642
```

RESULT 10

```
BJ462718
LOCUS        BJ462718                539 bp    mRNA      linear   EST 23-MAY-2002
DEFINITION   BJ462718 K. Sato unpublished cDNA library, cv. Haruna Nijo
             germination shoots Hordeum vulgare subsp. vulgare cDNA clone
             bags24b06 5', mRNA sequence.
ACCESSION    BJ462718
VERSION      BJ462718.1  GI:21141225
KEYWORDS     EST.
SOURCE       Hordeum vulgare subsp. vulgare.
ORGANISM     Hordeum vulgare subsp. vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
             ; Triticeae; Hordeum.
REFERENCE    1  (bases 1 to 539)
AUTHORS      Sato,K., Saisho,D. and Takeda,K.
TITLE        Barley EST sequencing project in NIG and Okayama Univ
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="bags24b06"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo germination shoots"
                     /tissue_type="shoots"
                     /dev_stage="germination"
BASE COUNT    147 a    115 c    139 g    138 t
ORIGIN        1..539
```

Alignment Scores:
```
Pred. No.:                          539
Score:                   2.27e-25    Matches:          66
Percent Similarity:      278.00      Conservative:     27
Best Local Similarity:   60.39%      Mismatches:       53
Query Match:             42.86%      Indels:            8
DB:                      18.31%      Gaps:              4
                         13
```

US-09-308-397-2 (1-306) x BJ462718 (1-539)

```
BASE COUNT      188 a    204 c    213 g    244 t    1 others
ORIGIN                 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

Alignment Scores:
Pred. No.:              1.42e-24
Score:                  274.50
Percent Similarity:     56.25%
Best Local Similarity:  39.20%
Query Match:            18.08%
DB:                     12

Length:       850
Matches:      69
Conservative: 30
Mismatches:   66
Indels:       11
Gaps:         4

US-09-308-397-2 (1-306) x BG444706 (1-850)

Qy      4 ThrAlaPheLeuPheGluAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyr 23
            :::::||||||::|||:::||:::|||||||::::|:|::|||:::|||||||||||:::
Db    222 TCCGCCCTTCCCTTTCCCGGTCAGGGGCGGCTCAGCCCGTTGGAATGGGAAAAGAA---GCT 278

Qy     24 AspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp 43
            ||||||||:::||:|||||::||||||||||||||  ::|||||||||||||||||
Db    279 CAGGCTGTGCCTGCTGCTGCAGAATTGTATAAGAAGGCAAATAATATTAGGTTTTTGAT 338

Qy     44 LeuArgTyrLeu--IleAspThrGluGluGlnLysLeuAsnGlnThrArgTyrThrGln 62
            |||           ::|||||||||||||||:::||   |||  :::|||:::
Db    339 CTTCTAGATCTTTGCATCAATGGTCCCAAGGAAAAGTTAGATTCAACTGTCATAAGCCAG 398

Qy     63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeuGlyLysValGlyTyrGln 82
            :||||||||||:||||||||||||||:||::|||::|
Db    399 CCTGCTATTTATGTGCACAAGTCTAGCTGCCGTGGAGCTGCTTCGTGCCGTGATGGAGGC 458

Qy     83 Pro----------AspMetValAlaGlyLeuSerLeuGluTyrSerAla 96
            :::            ||| ||||||||:||||::||:|||
Db    459 CCGGACAGATTATTGATTCTGATGTCACGTGGGTCTCAGCTTGGGATAATACTGCA 518

Qy     97 LeuValAlaSerGlyLeuAlaLeuAspPheGluAlaValAlaLeuValAlaLeuValAlaLysArgGly 116
            :::|||  ||| :|||||||||  :::|||:::|||
Db    519 CTAGCTTTTGCTGGGGCTTTAGCTTTGAGGATGGACTCAAGCTTGTCAAACTAAGGAGGA 578

Qy    117 AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
            |||||||||||||||||||::||||||:||||::|||||||||||||||||
Db    579 GAAGCCATGCAGGAGGCCGCAGATGCTGCTAAAAGTGCTGCTAAAGTCAGTATCATAGGACTT 638

Qy    137 ProValGluValIleGluGluAlaCysGlnLysAlaSerGluValGly 153
            |||||||||||:|||||||| |||::|::|||
Db    639 GATTCAGAGAAGGTTCAGCAATTGTGTGATGCTGCCACCAAGAATTGATGGAGCTGATAA 698

Qy    154 ValThrProAlaAsnTyrAsnThrProAlaGlnIleValIleAlaGly 169
            :::||||||||:||||||||||||||||||||||||||||||||||||
Db    699 GTGCAGATTGCGAATTACTTGTGTCCTGGAAATATGCAGTATCTGGGG 746

RESULT 12
BJ462722
LOCUS       BJ462722               513 bp    mRNA        linear    EST 23-MAY-2002
DEFINITION  BJ462722 K. Sato unpublished cDNA library, cv. Haruna Nijo
            germination shoots Hordeum vulgare subsp. vulgare cDNA clone
            bags24b17 5', mRNA sequence.
ACCESSION   BJ462722
VERSION     BJ462722.1  GI:21141229
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..513
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo germination shoots"
                     /clone="bags24b17"
                     /tissue_type="shoots"
                     /dev_stage="germination"
BASE COUNT      145 a    102 c    134 g    131 t
ORIGIN

Alignment Scores:
Pred. No.:              6.88e-25
Score:                  274.00
Percent Similarity:     59.74%
Best Local Similarity:  42.21%
Query Match:            18.05%
DB:                     13

Length:       513
Matches:      65
Conservative: 27
Mismatches:   54
Indels:       8
Gaps:         4

US-09-308-397-2 (1-306) x BJ462722 (1-513)

Qy      4 ThrAlaPheLeuPheGluAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyr 23
            :::::||||::|:|||:::||:::|||||||::::|:|::|:::|||||||||||:::
Db     17 ACCGNGTTCCTTTTCCCGGTCAGGGGCGGCTCAGCCCGTTGGAATGGGTAAAGAAGCTCTT 76

Qy     24 AspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp 43
            ||:||||||:::||:|||||::||||||||||||||  ::|||||||||||||||:::
Db     77 AAT--GTTCGAGCAGCTGCAGAACTATTGATAAGGCAAATGATATATCTTGGCTACGAC 133

Qy     44 LeuArgTyrLeu--IleAspThrGluGluGlnLysLeuAsnGlnThrArgTyrThrGln 62
            |||           ::|||||||||||:::||   |||  :::|||:::
Db    134 TTGCTGAATTTTGCATCGATGGACCAAAAGAAAAGCTGAACTGAACAGTGATCAGTCAG 193

Qy     63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgLeuLeu------GlnGluLysGly 80
            :||||||||||:|||||||||||||||:::||::|||        :::|||||||||
Db    194 CCAGCTATATATGTTACCACCCTTGCAGCTGTAGAAGTGTTACGTGCACGGGAAGACGGT 253

Qy     81 Tyr----------GlnProAspMetValAlaGlyLeuSerLeuGluTyrSerAla 96
            :::            ||| |||||:||||||:||||::||:|||
Db    254 CAACTGTATTAACTCTGTAGATGTCACATGTGGTCTCAGCTTGGGAGAATATACCGCCG 313

Qy     97 LeuValAlaSerGlyLeuAlaLeuAspPheGluAlaValAlaLeuValAlaLeuValAlaLysArgGly 116
            :::|||  ::||| |||:::|||  :::|||:::|||:::
Db    314 CTTGCATTTGCTGGTGCCTTAGCTTTGAGGATGTCTGAAGCTTGTCAAGCTAGAGGA 373

Qy    117 AlaTyrMetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
            |||||||||||||||||||::||||||:||||::||||||||||||:::||:::|||:::
Db    374 GAAGCCATGCAGGAGGCCGCAGATGCTGCCAATAGTGCGGATGGCTAGTGTGATTGGTCTA 433

Qy    137 ProValGluValIleGluGluAlaCysGlnLysAlaSerGlu 150
            |||||||||||:||||:||||| :|::|||:::
Db    434 GATTCAGAAAAGGTGCAACAATTATGCGATGCTGCAAATGAG 475

RESULT 13
AY108033
LOCUS       AY108033               892 bp    mRNA        linear    HTC 25-MAY-2002
DEFINITION  Zea mays PCO095766 mRNA sequence.
ACCESSION   AY108033
VERSION     AY108033.1  GI:21211111
KEYWORDS    HTC.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
```

```
JOURNAL      Overgo Probes
REFERENCE    Unpublished (2002)
AUTHORS      Coe,E.C.
TITLE        Direct Submission
JOURNAL      Submitted (25-APR-2002) Maize Mapping Project, University of
             Missouri, Columbia, MO 65211, USA
FEATURES          Location/Qualifiers
     source       1..892
                  /organism="Zea mays"
                  /db_xref="MaizeDB:635450"
                  /db_xref="taxon:4577"
                  /clone="PC009576"
                  /clone_lib="Maize Mapping Project/DuPont Cornsensus
                  Library"
                  /note="this sequence is part of a project of EST
                  assemblies resulting from the application of public
                  contigs to seed DuPont contigs; this resource was
                  assembled by DuPont as part of a collaboration for the
                  overgo addressing of BACs in conjunction with the Maize
                  Mapping Project"

BASE COUNT     191 a     261 c     238 g     202 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                      5.92e-24              Length:           892
                            270.00                Matches:          66
Percent Similarity:         56.65%                Conservative:     32
Best Local Similarity:      38.15%                Mismatches:       57
Query Match:                17.79%                Indels:           18
DB:                         11                    Gaps:             5

US-09-308-397-2 (1-306) x AY108033 (1-892)

Qy      4 ThrAlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyr  23
             |||||||||||  ||||||||||||||  :::||||||||||||||||||
Db    377 ACCGCTTCCTTCCTTCCCCGGCCAGGTGCCCAAACCGTTGGAATGGGGGGCAGAA---GCT 433

Qy     24 AspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp  43
             :::  ||||||||||||  ||||||||||||||||||  :::
Db    434 CAGAGTGTTCCAGCAGCTACCAAATTGTTCAACCAGCAAATGAAATCCTTGGATATGAC  493

Qy     44 LeuArgTyrLeu---IleAspThrGluGluAspLysLeuLeuAsnGlnThrArgTyrThrGln  62
             |||||||||||        ::::          |||||||   :::||
Db    494 TTGCTGGATCTTGCACCAATGGACACCAAAAGAAAAGCTTGACTCAACAATGATCAGTCAG 553

Qy     63 ProAlaIleLeuAlaThrSerValAlaIleTyrArgValLeuGlnGluLys---GlyTyr  81
             :::||||||||||||  |||  :::||  ||||||||
Db    554 CCAGCTATATATGTCACCAGCCTTGCAGCAGTAGAGTTACACGGTGCACGCGATGGAGGC 613

Qy     82 GlnPro----------AspMetValAlaGlyLeuSerLeuGlyTyrGluTyrSerAla  96
             |||                ||||||  ||||||||||||||||
Db    614 CAAGATGTGATTAACTCTGTAGATGTTACATGTGGTCCAGTTTGGGAGTATACCGCT 673

Qy     97 LeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLeuLysArgGly 116
             ||||||||  :::  |||  ||  ||||||||||
Db    674 CTTGCATTGCCGGTGCTTTAGCTTTGAGGATGGACTGAAGCTTGTCAAGCTAAGAGGA 733

Qy    117 AlaTyrMetGluGluValAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThr 136
             ||||||||||||  :::      |||  |||
Db    734 GAAGCTATGCAGGATGCTTCAGATGTCCAATAGTGCGATGGTTAGTGTGATTGGTCTG 793

Qy    137 ProValGluValIleGluGluAlaCysGlnLysAlaSerGlu----------  150
             ||||||||||||||||||  :::::
Db    794 GATTCAGAAAAGGTGCAAGAACTATGTGATGCTAATGACGAGTAGATGAAAACGAT 853

Qy    151 ----------LeuGlyValValThrProAlaAsnTyr 159
                         :::  ||||||
Db    854 AGAGTTCAAATAGCAAACTTTCTGTGCCCTGGAATTAT 892

RESULT 14
AF029428/c
LOCUS       AF029428            1237 bp    DNA      linear       GSS 29-AUG-2000
```

```
DEFINITION   AF029428 Salmonella typhimurium LT2, Lambda DASH II Salmonella
             typhimurium genomic clone 1257-T7, DNA sequence.
ACCESSION    AF029428
VERSION      AF029428.1  GI:2570958
KEYWORDS     GSS.
SOURCE       Salmonella typhimurium.
  ORGANISM   Salmonella typhimurium
             Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
             Salmonella.
REFERENCE    1 (bases 1 to 1237)
  AUTHORS    Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
  TITLE      Sample sequencing of a Salmonella typhimurium LT2 lambda library:
             comparison to the Escherichia coli K12 genome
  JOURNAL    FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
   MEDLINE   99243757
COMMENT      Contact: McClelland M
             Molecular Biology
             Sidney Kimmel Cancer Center
             3099 Science Park Road, San Diego, CA 92121, USA
             Email: mcclelland@lifsci.sdsu.edu
             Class: shotgun.
FEATURES          Location/Qualifiers
     source       1..1237
                  /organism="Salmonella typhimurium"
                  /strain="LT2"
                  /db_xref="taxon:602"
                  /clone="1257-T7"
                  /clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
                  /note="Vector: Lambda DASH II; sequenced using Li-Cor
                  sequencer"

BASE COUNT     293 a     345 c     304 g     281 t
ORIGIN

Alignment Scores:
Pred. No.:                                        Length:          1237
Score:                      1.04e-23              Matches:          62
Percent Similarity:         270.00                Conservative:     21
Best Local Similarity:      65.87%                Mismatches:       37
Query Match:                49.21%                Indels:           6
DB:                         17.79%                Gaps:             3
                            17

US-09-308-397-2 (1-306) x AF029428 (1-1237)

Qy      1 MetThrLysThrArgPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArg  20
             ||||||||||||||||||||||||||||||||||||||||||||
Db    398 ATGACGCATTTGCATTTGTGTTCCCCGGTCAGGGTTCAGAGGCGTTGGGGATGTTGGCC 339

Qy     21 AspPheTyrAspGlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeu  40
             :::  ||||||||||||||  ||||||||||||||||||||
Db    338 GAGATGGGGGCAAATTACCCTATCGTAGAAGAAACGTTGCTGAAGCTTCTCGGCGTCTG 279

Qy     41 GlyTyrAspLeuArgTyrLeuIleAspThrGluGlu---------AspLysLeuAsnGln  57
             |||||||                  ::::
Db    278 GGATATGATCTGTGGGCGCTC----ACCCAGCAAGGTCCAGCGGAAGAACTGAATAAA 225

Qy     58 ThrArgTyrThrGlnProAlaIleLeuAlaThrSerValAlaIleTyrArgValLeuGln  77
             ||||||||||||
Db    224 ACCTRGCAGCACGCCAGCCGCGGTTATTAACCGCTTCCGNTSCGMTWTGGCGCRTTTGGCAG 165

Qy     78 GluLys---GlyTyrGlnProAspMetValAlaGlyLeuSerLeuGlyTyrGluTyrSerAla  96
             :::::  :::  ||||||||||||||||||||||
Db    164 CAGCAGGGCGGGTAAAATGCCTGCGTWAATGCAGGTCACAGCCTGGGCGGAATATTCCGCG 105

Qy     97 LeuValAlaSerGlyAlaLeuAspPheGluAspAlaValAlaLeuValAlaLeuLysArgGly 116
             |||||  :::  :::  ||||||||||
Db    104 CTRGTTTGMGCTGCGTCATCAACTTTCGTGATGCCGGTTCGTCGTGGTGGAAAATGCGCGGT  45

Qy    117 AlaTyrMetGluGluAla 122
             :::  |||
Db     44 AAATTCATGCAGGAAGCG  27

RESULT 15
BE237947
```

```
LOCUS       BE237947                599 bp    mRNA    linear   EST 11-JUL-2000
DEFINITION  894039B1.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BE237947
VERSION     BE237947.1  GI:9027907
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Elizabeth H. Harris
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000, USA
            Tel: 919 613 8164
            Fax: 919 613 8177
            Email: chlamy@duke.edu.
FEATURES             Location/Qualifiers
     source          1..599
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT       96 a    193 c    214 g     96 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                 4.01e-24        Length:        599
Percent Similarity:    269.00          Matches:        61
Best Local Similarity: 60.14%          Conservative:   28
Query Match:           41.22%          Mismatches:     53
DB:                    17.72%          Indels:          6
                       10              Gaps:            3

US-09-308-397-2 (1-306) x BE237947 (1-599)

Qy    5 AlaPheLeuPheAlaGlyGlnGlyAlaGlnTyrLeuGlyMetGlyArgAspPheTyrAsp  24
         | || |||| |||  ||||||||||||||||||||||||||   ::|||
Db  155 GCGGCATTGTTCCCGGCCAGGGTGCCCAGGTGCCCAGAGCGTCGGCATGGCCAGGACTTGTGGGCG  214

Qy   25 GlnTyrProIleValLysGluThrIleAspArgAlaSerGlnValLeuGlyTyrAsp---  43
         :||||:|||||||||||:::||||||||||||:|||||||||||:||||||||:||
Db  215 ACCGTGCGCAGGGCCAAGGAAGGATGTTCGACAAGGGCGTCTGAGATCCTGGGCATACGACCTG  274

Qy   44 LeuArgTyrLeuIleAspThrGluGluLysLeuAsnGlnThrArgTyrThrGlnPro  63
         ||||:|     |||:||||:::||||||   ||  ||||||
Db  275 CTGAAGGTGTGCGTGGAGGGGCCCCAAGGAGGAGAAGCTGGACAGCACCGCCGTGAGTCAGCCC  334

Qy   64 AlaIle---LeuAlaThrSerValAlaIleTyrArgLeuLeuGlnGluLysGlyTyrGln  82
         ::|||:    :|||:|||      :::|||
Db  335 GCTATCTACGTGCCTCGCTGGCCGTGGCCGTGGAGAAGCTGCGCGCGGACGAGGGCCAGGCC  394

Qy   83 Pro--------AspMetValAlaGlyLeuAlaGlyLeuSerLeuGlyGluGlyTyrSerAlaLeuVal  98
         ||:::         |||||||||||||||||:||||||::::|||::|||||
Db  395 GCCATTGACCGCCATTGACGTGGCGTGCGGCCTGCTGCGGCGGCGAGTACACGGGCCCTGGCC  454

Qy   99 AlaSerGlyAlaLeuAspPheGluAlaValAlaLeuAlaLeuAlaLeuLysArgGlyAlaTyr  118
         :::|||| :::     ||||||     :::     ||||:     ::|
Db  455 TTCGCTGGTGCCATGAGCTTCGAGGACGGCGTCGGCGGGCTGCGCGGCGGGCGGAGTCC  514

Qy  119 MetGluGluAlaAlaProAlaAspSerGlyLysMetValAlaValLeuAsnThrProVal  138
         ||:::|||||||:|:::
Db  515 ATGCAGGCGGCTGCTGACGGCGCAGCCAGCGACGCATGGTGTCCGTCATTGGGCTGGACAGC  574

Qy  139 GluValIleGluGluAlaCysGln  146
         ||:::
Db  575 GCCAAGGTGGCGGAGCTGTGCAAG  598

Search completed: June 11, 2003, 21:24:41
Job time : 1437 secs
```